

Review

Genomic Selection for Forest Tree Improvement: Methods, Achievements and Perspectives

Vadim G. Lebedev ^{1,*}, Tatyana N. Lebedeva ², Aleksey I. Chernodubov ³ and Konstantin A. Shestibratov ^{1,3}

¹ Branch of the Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry of the Russian Academy of Sciences, Prospekt Nauki 6, 142290 Pushchino, Moscow Region, Russia; schestibratov.k@yandex.ru

² Institute of Physicochemical and Biological Problems in Soil Science of the Russian Academy of Sciences, Institutskaya Str. 2, 142290 Pushchino, Moscow Region, Russia; tanyaniko@mail.ru

³ Department of Forest Crops, Selection and Forest Reclamation, Voronezh State University of Forestry and Technologies Named after G.F. Morozov, 8 Timiryazeva Str. 394087 Voronezh, Russia; chernodubov2010@yandex.ru

* Correspondence: vglebedev@mail.ru

Received: 25 October 2020; Accepted: 8 November 2020; Published: 11 November 2020



Abstract: The breeding of forest trees is only a few decades old, and is a much more complicated, longer, and expensive endeavor than the breeding of agricultural crops. One breeding cycle for forest trees can take 20–30 years. Recent advances in genomics and molecular biology have revolutionized traditional plant breeding based on visual phenotype assessment: the development of different types of molecular markers has made genotype selection possible. Marker-assisted breeding can significantly accelerate the breeding process, but this method has not been shown to be effective for selection of complex traits on forest trees. This new method of genomic selection is based on the analysis of all effects of quantitative trait loci (QTLs) using a large number of molecular markers distributed throughout the genome, which makes it possible to assess the genomic estimated breeding value (GEBV) of an individual. This approach is expected to be much more efficient for forest tree improvement than traditional breeding. Here, we review the current state of the art in the application of genomic selection in forest tree breeding and discuss different methods of genotyping and phenotyping. We also compare the accuracies of genomic prediction models and highlight the importance of a prior cost-benefit analysis before implementing genomic selection. Perspectives for the further development of this approach in forest breeding are also discussed: expanding the range of species and the list of valuable traits, the application of high-throughput phenotyping methods, and the possibility of using epigenetic variance to improve of forest trees.

Keywords: forest tree breeding; genomic selection; molecular markers; high-throughput phenotyping; epigenetics; genotyping; genomic prediction models; quantitative trait locus; breeding cycle

1. Forest Tree Breeding

1.1. Traditional Breeding

In contrast to agricultural crops with their breeding history of centuries or even millennia, the history of forest tree breeding is relatively recent. It was only in the late 1950s in the USA and Europe that the increasing demand for wood for construction, fuel, and paper production, and the reduction in natural forests generated the need for forest tree domestication using modern breeding methods [1]. The breeding of woody plants is much more difficult, longer, and expensive than that of agricultural crops. Trees have a long juvenile period (late age flowering and seed production) and a large physical size; their progeny testing requires large areas and long-term observations.

Some important forest tree traits, e.g., tolerance to biotic and abiotic stresses, are difficult to assess in the field. For forest trees in general and coniferous trees in particular, the stages of reproduction and progeny testing may take up 15 years each [2]. In the USA, one cycle of pine species improvement with classical methods can take about 30 years, including breeding (at least 10 years), field testing (at least 8 years), and propagation (at least 8 years) [3]. Even for the fast-growing *Eucalyptus globulus* Labill., improvement programs take 12 years and are quite expensive [4]. Furthermore, it is nearly impossible to assess the most economically important traits (wood quality, growth rate) in adult trees (aged 30–50 years), and therefore, selection has to be based on measurements made in young trees (aged 5–15 years). Normally, selection is made at about one-third of the rotation age, e.g., 2–3 years for *Eucalyptus* and 6–12 years for *Pinus* spp. [5]. However, it is known that tree traits at juvenile and mature ages poorly correlate with each other [6]. Moreover, most of the complex tree traits (growth, stem form, and branching) have low or moderate heritability, which limits the selection response and thereby hampers the genetic gain [7]. Unlike annual crops and fruit trees, forest trees were practically not subject to domestication, which would have reduced their genetic diversity, and they actually represent wild populations. Their nearly absent pedigree makes it difficult to estimate their breeding values. Moreover, possible errors in the determined progeny relationships among open-pollinated species (pines, spruces) may lead to lower accuracy of estimated breeding values (EBVs) [8]. EBV is the genetic merit of individual plants that refers to its value in a breeding program for a particular trait. Thus, forest trees have a slow accumulation of genetic gain per unit time.

Due to the above features, many forest tree breeding programs are now only in the first or second cycle of breeding, and only the most advanced programs are in their third or fourth cycle [2]. For example, it took 55 years to complete three breeding cycles in a loblolly pine (*Pinus taeda* L.) breeding program in the USA [9]. The classical approach based on phenotypic assessment helped improve the tree genotypes and reduce the rotation time, yet the speed of breeding was absolutely unsatisfactory. Advanced methods of breeding (e.g., top-grafting) and propagation (e.g., somatic embryogenesis) can shorten the stages in coniferous trees from 10 to 4 years, and from 8 to 1 year, respectively [3], but the duration of field testing still remains unchanged. Long breeding cycles of forest trees hinder the growing demand for new valuable genotypes for forest plantations that are increasingly expanding around the world. Ongoing global climate change also requires the accelerated development of new genotypes of forest trees, especially stress-resistant ones. Therefore, breeders are looking for new ways to improve the efficiency of forest breeding, i.e., to accelerate the selection of valuable genotypes.

1.2. Marker-Assisted Selection

There were attempts to speed up the selection process by using genotypic rather than phenotypic evaluation of genotypes, i.e., using molecular markers. The first use of molecular markers in traditional phenotypic selection led to the development of marker-assisted selection (MAS). This method has been used in plant breeding since the 1990s [10]. The principle of MAS consists of exploiting linkage disequilibrium (LD) between markers and quantitative trait loci (QTLs), i.e., non-random association between marker and QTL alleles [5]. MAS is most effective for simple traits controlled by a few QTLs, each responsible for a relatively large part of the overall phenotypic variability. It was for simply inherited traits, e.g., disease resistance, that some success was achieved with MAS in forest trees [11]. However, the most economically important tree traits (wood properties, trunk straightness, growth rate) have quantitative, complex inheritance patterns, and the use MAS was much less effective for them. Such traits are linked with multiple QTLs, each having a small effect, and MAS is not designed to track a large number of loci. For example, as shown on two populations of *Pinus taeda*, QTLs explained as little as 5.3% to 15.7% of the phenotypic variance in several wood traits, and these populations, in turn, contained only part of the genetic variability present in the entire population of loblolly pine [12]. Errors in defining a stringent statistical threshold for declaring a significant effect of a particular QTL can lead to a high percentage of false-positive or false-negative results. Being undomesticated, forest trees are also characterized by low levels of LD [13]. Finally, successful use of MAS in forest

trees was hindered by strong QTL-environment and QTL-genetic background interactions, and the fluctuation of alleles frequency over generations [14,15]. On the whole, it was shown that QTLs do not explain enough genetic variations to enable any effective implementation of MAS for complex traits in forest trees [11].

1.3. Genomic Selection

About 20 years ago, Meuwissen et al. [16] proposed an alternative approach based on analysis of all QTL effects, regardless of their significance. This technology began developing after the costs of genotyping had dropped due to advances in next generation sequencing (NGS) and the use of large amounts of SNP markers, as well as the development of statistical methods for large dataset analysis. Genomic selection (GS) first became widely used in animal breeding, and then in crop breeding. GS proved to be thrice as efficient as MAS in maize and twice as efficient as MAS in wheat [17]. GS is also based on molecular markers, but, unlike MAS, preliminary information about phenotype-marker linkage, localizations of QTLs in the genome and their relative effect on the phenotype are not prerequisite for GS [18]. In simple words, the key difference is as follows: MAS uses few markers linked with large-effect QTLs, whereas GS uses lots of markers linked with minor-effect QTLs. The use of high-density markers is one of the fundamental features of GS: each locus of a trait has likelihood to be in LD with at least one marker locus in the entire target population [10]. To date, GS has demonstrated promising results in plants, with prediction accuracy higher than that of MAS [19].

Compared to phenotypic selection-based traditional breeding, GS can significantly shorten the breeding cycle because the marker-based assessment of genotypes can be done at a very early stage, when DNAs can be isolated without harming the plants. This eliminates the need for lengthy and expensive field testing of progeny, otherwise necessary for phenotypic evaluation. The possibility of early assessment is especially important for traits that manifest at late stages of development (particularly in long-living species) or are difficult to assess with an adequate accuracy (e.g., pest- and disease-resistance). Ultimately, this will significantly increase genetic gain per unit time [20,21].

GS is relatively seldom used in forest tree breeding, in contrast to crop breeding. The large and poorly studied genomes of trees, conifers in particular, make their genotyping difficult. However, GS can significantly reduce the breeding cycle, which is particularly important for slow-growing boreal coniferous species with a breeding cycle of 20 to 30 years. For example, the use of GS reduced the field testing stage to several months [3]. As a result, GS nearly halved the breeding cycle of loblolly pine: it made phenotype-based assessment unnecessary and allowed marker-based genotyping as soon as seeds were available. A similar effect was obtained in deciduous trees: GS reduced the breeding cycle of eucalyptus from 10 to 5 years [11]. One of the first GS studies already showed that, for eucalyptus species, a 50% shortening of the breeding cycle would increase efficiency gains by 50–100%, and a 75% shortening would increase the efficiency by 200–300% [22]. Stimulation of early flowering in individual plants can shorten the breeding cycle even more: eucalyptus breeding programs achieve flowering at 1–4 years compared to 4–8 years in natural conditions, and in *Pinus taeda*, by grafting, flowering can be achieved at 3 years [23]. In addition, GS allows simultaneous and early selection for several traits among a large number of individuals—an impossible task for conventional tree breeding now mostly based on tandem selection [24]. Thus, GS can provide a higher genetic gain per unit time, a faster succession of generations in breeding programs, and a faster creation of genotypes better adapted to environmental changes (the global climate change, the spread of diseases), and can better meet the industrial demand for wood. Thus, although GS can potentially improve the efficiency of tree breeding programs, the outcomes would strongly depend on species and traits of interest. Therefore, cost-benefit analysis is essential.

Due to the low degree of forest tree domestication, long traditional breeding cycles, and a large genetic variability of almost any trait, GS will probably be much more efficient in forest breeding than the traditional approach and, perhaps, even more efficient than in crops and animals [2]. Since the first simulation [23,25] and empirical studies [22,26] in the early 2010s, GS has attracted considerable

interest among forest breeders around the world: about 60 studies on this topic have been conducted to date (Tables 1 and 2). As shown in Table 2, more than 80% of them were performed in trees from the genera *Eucalyptus*, *Picea*, and *Pinus*, the main tree plantation species in the world. In the last 3–4 years alone, the list was expanded by adding less common forest species, such as Douglas-fir, Japanese cedar, rubber tree, etc. As in traditional breeding, phenotype was mainly assessed based on growth and wood traits (mainly the physical ones), less often, based on tree architecture and pulp yield, and very rarely, based on species-specific traits, e.g., the content of essential oils and their components in *Eucalyptus polybractea* R.T.Baker [8] or rubber production in *Hevea brasiliensis* (Willd. ex A. Juss.) Müll. Arg. [27]. More recently, there appeared studies on GS for resistance to stresses, such as diseases [28,29], pests [30,31], and drought [32].

Table 1. Summary of the published simulation studies on GS in forest trees.

Species	Population Size	Effective Population Size	Heritability of Trait	Number of QTL	Marker Density, Marker/cM	Prediction Model ¹	Scenario ²	Reference
forest trees	200–8000	10–1000	0.2–0.6	1–200	1–20	BLUP	PS GS	[23]
<i>Cryptomeria japonica</i> (Thunb. ex L.f.) D.Don	2000	-	0.1–0.5	50 or 150	0.5–5	linear	PS GS PS + GS	[25]
<i>Eucalyptus</i>	-	100	0.1 or 0.6	44	0.1	BLASSO	PS GS GS + dominance	[33]
<i>Pinus taeda</i>	1000	-	0.25	30 or 1000	8.33	BRR, Bayes A, Bayes B, BLASSO	GS GS + dominance	[34]
conifers	800–3000	-	0.2 or 0.5	350	4–53.3	Bayes C	PS PS + clonal archive GS GS + top grafting	[35]
forest trees	2050	5–25	0.2 or 0.6	100–200	1–10	ABLUP, GBLUP	GS	[36]
<i>Hevea brasiliensis</i> (Willd. ex A. Juss.) Müll. Arg.	100–5000	-	-	-	-	linear	PS PS + GS	[27]

¹ BLUP—best linear unbiased prediction, BLASSO—Bayesian least absolute shrinkage and selection operator, BRR—Bayesian ridge regression, ABLUP—additive BLUP, GBLUP—genomic BLUP. ² PS—phenotypic selection, GS—genomic selection.

Table 2. Summary of the published experimental studies on GS in forest trees.

Species	Population Structure ¹	Genotyping Technology	Marker Type	Number of Markers	Traits ²	Phenotyping Age, Years ³	Prediction Model ⁴	Breeding Cycle in PS/GS, Years ⁵	Reference
<i>Castanea dentata</i> (Marshall) Borkh.	1230 (HS)	GBS	SNP	71,507	disease resistance	5–16	ABLUP, HBLUP, Bayes C		[28]
<i>Cryptomeria japonica</i>	476 (WP)	Axiom array	SNP	32,036	growth, physWP, male fecundity	10	GBLUP, Bayes B, RF		[37]
<i>C. japonica</i>	547 (FS), 29 (HS)	Axiom array	SNP	3034	growth, physWP	18	GBLUP, RF		[13]

Table 2. Cont.

Species	Population Structure ¹	Genotyping Technology	Marker Type	Number of Markers	Traits ²	Phenotyping Age, Years ³	Prediction Model ⁴	Breeding Cycle in PS/GS, Years ⁵	Reference
<i>Eucalyptus benthamii</i> Maiden and Cambage <i>E. pellita</i> F. Muell.	505 (HS) 732 (HS)	EUChip60K	SNP	13,787 19,506	growth	56 months 42 months	ABLUP, GBLUP, Bayes A, Bayes B, Bayes C π , BLASSO, BRR		[38]
<i>E. cladocalyx</i> F. Muell.	480 (49 HS)	EUChip60K	SNP	3879	growth, physWP, treeArc, flowering intensity	17 or 18	Bayes A, Bayes B, Bayes C, BRR		[39]
<i>E. globulus</i>	310 (40 FS, 13 HS)	EUChip60K	SNP	~12,000	growth	5–11	GBLUP, Bayes B, Bayes C, BLASSO		[4]
<i>E. globulus</i>	647 (62 FS, 3 HS)	EUChip60K	SNP	14,442	growth, treeArc	4	RR-BLUP, RR-BLUP-B, Bayes A, Bayes B, BLASSO, PCR, S-PCR		[6]
<i>E. globulus</i>	646 (FS, HS)	EUChip60K	SNP	14,422	growth, physWP, treeArc	6	Bayes A, Bayes B, Bayes C, BLASSO, BRR		[40]
<i>E. grandis</i> W. Hill \times <i>E.</i> <i>urophylla</i> S.T. Blake <i>E. grandis</i> , <i>E. urophylla</i> , <i>E. globulus</i> and hybrids	738 (43 FS) 920 (75 FS)	DArT Pty array	DArT	3129 3564	growth, physWP, pulp yield	3 3 or 3.7	RR-BLUP	18/9	[22]
<i>E. grandis</i>	187 (131 HS)	DArT Pty array	DArT	7680	growth	5	ABLUP, HBLUP		[41]
<i>E. grandis</i> \times <i>E. urophylla</i>	768 (37 FS)	EUChip60K	SNP	24,806	growth, physWP, pulp yield	3	GBLUP (+ D)	18/14 or 9	[42]
<i>E. grandis</i>	187 (131 HS)	DArT Pty array	DArT	2816	growth, treeArc	5	ABLUP, HBLUP (+ MT)		[43]
<i>E. grandis</i> , <i>E. grandis</i> \times <i>E. camaldulensis</i> Dehnh.	995 (69 FS)	sequence capture, EUChip60K	SNP	55,722, 40,932	growth, physWP, chemWP	24 and 36 months	RR-BLUP, BayesB		[44]
<i>E. grandis</i> \times <i>E. urophylla</i>	999 (45 FS)	EUChip60K	SNP	33,398	growth, physWP, chemWP	5	ABLUP, GBLUP, HBLUP		[45]

Table 2. Cont.

Species	Population Structure ¹	Genotyping Technology	Marker Type	Number of Markers	Traits ²	Phenotyping Age, Years ³	Prediction Model ⁴	Breeding Cycle in PS/GS, Years ⁵	Reference
<i>E. grandis</i>	1548 (FS, HS)	EUChip60K	SNP	15,040	growth, physWP, chemWP	7	ABLUP, GBLUP	8/4	[46]
<i>E. nitens</i> (H. Deane and Maiden) Maiden	691 (72 HS)	EUChip60K	SNP	12,236	growth, physWP	6 or 7	GBLUP, BLUP		[47]
<i>E. pellita</i>	468 (28 HS)	marker panel	SNP	2023	growth, pulp yield	54 months	ABLUP, GBLUP, HBLUP		[48]
<i>E. polybractea</i> R.T. Baker	480 (40 HS)	whole genome sequencing	SNP	10,000, 97,000 and 502,000	growth, terpene, and essential oil content	1 and 2	ABLUP, GBLUP, Bayes B		[8]
<i>E. robusta</i> Sm.	415 (WP)	DArTseq	SNP	2919	growth, chemWP	49 months or 20 years	GBLUP (+ MT)		[49]
<i>E. urophylla</i> × <i>E. grandis</i>	1130 (69 FS)	DArTseq	SNP	3303	growth	32 months	PBLUP (+ D, E)		[50]
<i>E. urophylla</i> × <i>E. grandis</i>	958 (338 FS)	EUChip60K	SNP	41,304	growth, physWP, pulp yield	3 and 6	ABLUP, GBLUP, RR-BLUP, BLASSO, RKHS		[51]
<i>E. urophylla</i> × <i>E. grandis</i>	958 (338 FS)	EUChip60K	SNP	41,304	growth, physWP, pulp yield	3 and 6	GBLUP (+ D, E)		[52]
<i>E. urophylla</i> × <i>E. grandis</i>	1130 (69 FS)	DArTseq	SNP	3303	growth, chemWP, water use efficiency	55 months	GBLUP (+ MT)		[32]
<i>Fraxinus excelsior</i> L.	1250 (WP, HS)	whole genome sequencing	SNP	100–50,000	disease resistance	6	RR-BLUP		[29]
<i>Hevea brasiliensis</i>	435 (FS)	GBS	SNP	107,294	growth	1, 2, 3, and 4	GBLUP	10/3	[53]
<i>H. brasiliensis</i>	330 (FS)		SSR	332	rubber production	4	RR-BLUP (+ D), BLASSO, RKHS		[27]
<i>Picea abies</i> (L.) H. Karst.	1370 (128 FS)	sequence capture	SNP	116,765	growth, physWP	17 or 30	ABLUP, GBLUP, BLASSO, BRR, RKHS	25/15	[54]
<i>P. abies</i>	1370 (128 FS)	sequence capture	SNP	116,765	growth, physWP	17 or 30	ABLUP, GBLUP (+ D, E)		[55]

Table 2. Cont.

Species	Population Structure ¹	Genotyping Technology	Marker Type	Number of Markers	Traits ²	Phenotyping Age, Years ³	Prediction Model ⁴	Breeding Cycle in PS/GS, Years ⁵	Reference
<i>P. abies</i>	726 (40 FS)	Infinium iSelect array	SNP	5660	growth, physWP pest resistance	15 or 16 10 and 15	ABLUP, GBLUP, TGBLUP, Bayes C π , BRR (+ MT)		[30]
<i>P. abies</i>	484 (62 HS)	exome capture	SNP	130,269	physWP	19 or 22	ABLUP, GBLUP, RR-BLUP, Bayes B, RKHS		[56]
<i>P. glauca</i> (Moench) Voss,	1748 (FS, HS)	Infinium array PgLM3	SNP	6932	growth, physWP	17	ABLUP, RR-BLUP, BLASSO		[57]
<i>P. glauca</i>	1694 (214 HS)	Infinium array PgAS1	SNP	7338	growth, physWP	27	BLUP	30/10	[58]
<i>P. glauca</i> \times <i>P.</i> <i>engelmannii</i> Parry ex Engelm.	769 (25 HS)	GBS	SNP	34,570-50,803	growth	3, 6, 10, 15, 30, and 40	RR-BLUP, Bayes C π , GRR		[14]
<i>P. glauca</i> \times <i>P.</i> <i>engelmannii</i>	1126 (25 HS)	GBS	SNP	8868–62,618	growth, physWP	38	ABLUP, GBLUP, RR-BLUP, GRR (+ MT)		[15]
<i>P. glauca</i> \times <i>P.</i> <i>engelmannii</i>	1694 (214 HS)	Infinium array PgAS1	SNP	7338	growth, physWP	38	ABLUP, GBLUP (+ D, E)		[59]
<i>P. glauca</i>	1694 (214 HS)	Infinium array PgAS1	SNP	6716	growth, physWP	22	HBLUP		[60]
<i>P. glauca</i> \times <i>P.</i> <i>engelmannii</i>	1126 (25 HS)	GBS	SNP	~30,000	growth, physWP	38	ABLUP, GBLUP (+ D, E)		[61]
<i>P. glauca</i>	1516 (136 FS)	Infinium iSelect array	SNP	4148	growth, physWP, pest resistance	16–28	ABLUP, GBLUP, Bayes C π (+ D)		[31]
<i>P. glauca</i>	1513 (54 FS), 892 (HS)	Infinium array PgLM3	SNP	4092	growth, physWP,	19 or 20	ABLUP, GBLUP		[62]
<i>P. glauca</i> \times <i>P.</i> <i>engelmannii</i>	1126 (25 HS)	GBS	SNP	62,190	growth, physWP	38	ABLUP, RR-BLUP		[63]
<i>P. mariana</i> (Mill.) Britton, Sterns, and Poggenb.	734 (34 FS)	Infinium iSelect array	SNP	4993	growth, physWP	25	BLUP	28/9	[64]

Table 2. Cont.

Species	Population Structure ¹	Genotyping Technology	Marker Type	Number of Markers	Traits ²	Phenotyping Age, Years ³	Prediction Model ⁴	Breeding Cycle in PS/GS, Years ⁵	Reference
<i>P. sitchensis</i> (Bong.) Carr.	498 (3 FS)	RADseq	SNP	from ~2000–3000 to ~56,000	growth, timing of budburst	5 or 6	GBLUP		[65]
<i>Pinus contorta</i> Douglas ex Loudon	732 (FS), 697 (HS)	Axiom array	SNP	19,584	growth, physWP	10	ABLUP, GBLUP, Bayes B, Bayes C		[66]
<i>P. pinaster</i> Aiton	818 (35 HS)	Infinium array	SNP	4436	growth, treeArc	8 or 12	ABLUP, BLUP, BLASSO		[7]
<i>P. pinaster</i>	661 (FS, HS)	Infinium array	SNP	2500	growth, treeArc	6–15	GBLUP, BLASSO, BRR		[18]
<i>P. radiata</i> D. Don	1103	exome capture	SNP	67,168	treeArc, internal checking, external resin bleeding	7, 8, or 9	ABLUP, GBLUP	17/9	[67]
<i>P. sylvestris</i> L.	694 (183 FS)	GBS	SNP	8719	growth, physWP	10 and 30/30 and 36	ABLUP, GBLUP, BLASSO, BRR	36/18 and 21/11	[68]
<i>P. taeda</i>	951 (61 FS)	Infinium array	SNP	4853	growth, physWP, chemWP, treeArc, disease resistance (greenhouse), rooting ability (greenhouse)	4 or 6	RR-BLUP, RR-BLUP B, Bayes A, Bayes C π , BLASSO		[69]
<i>P. taeda</i>	926 (61 FS)	Infinium array	SNP	4825	growth	1, 2, 3, 4, and 6/3, 4, and 6	BLUP		[26]
<i>P. taeda</i>	149 (13 FS)	Infinium array	SNP	3406	growth, chemWP, treeArc	5	BLUP	15/7.5	[12]
<i>P. taeda</i>	165 (9 FS)	Infinium array	SNP	3461	growth	5	ABLUP, GBLUP		[9]
<i>P. taeda</i>	951 (61 FS)	Infinium array	SNP	4853	growth		PBLUP, GBLUP (+ D, E)		[70]
<i>P. taeda</i>	923 (71 FS)	Infinium array	SNP	7216	growth, disease resistance (greenhouse)	6	Bayes A, Bayes B, BLASSO, BRR		[34]

Table 2. Cont.

Species	Population Structure ¹	Genotyping Technology	Marker Type	Number of Markers	Traits ²	Phenotyping Age, Years ³	Prediction Model ⁴	Breeding Cycle in PS/GS, Years ⁵	Reference
<i>Populus deltoides</i> W. Bartram ex Marshall	473 (WP)	sequence capture	SNP	92,000	growth	1, 2, 3, 4 and 5 (field)1–15 weeks (greenhouse)	GBLUP		[71]
<i>Pseudotsuga menziesii</i> (Mirb.) Franco	1372 (37 FS)	exome capture	SNP	69,551	growth physWP	12 and 35 38	ABLUP, RR-BLUP, GRR		[72]
<i>P. menziesii</i>	1321 (37 FS)	exome capture	SNP	69,551	growth	12	ABLUP, RR-BLUP, Bayes B, GRR		[19]
<i>P. menziesii</i>	1321 (37 FS)	exome capture	SNP	56,454	growth, physWP	35 or 38	ABLUP, RR-BLUP		[63]
<i>Shorea platyclados</i> Sloot. ex Foxw.	356 (HS)	whole genome sequencing	SNP	5900	growth, physWP, treeArc	11	Bayes A, Bayes B, Bayes C, BL, BRR		[73]

¹ HS—half-sibs; FS—full-sibs; WP—wild population. ² physWP—physical wood properties; chemWP—chemical wood properties; treeArc—tree architecture. ³ x or y—different traits at different age; x and y—all traits at different age. ⁴ BLUP—best linear unbiased prediction; ABLUP—additive BLUP; GBLUP—genomic BLUP; TGBLUP—threshold genomic BLUP; HBLUP—single-step genomic BLUP; PBLUP—parent BLUP; RR-BLUP—ridge regression BLUP; BLASSO—Bayesian least absolute shrinkage and selection operator; BRR—Bayesian ridge regression; RKHS—reproducing kernel Hilbert space; GRR—generalized ridge regression; PCR—principal component regression; S-PCR—supervised PCR; D—dominance effect; E—epistatic effects; MT—multi-trait models. ⁵ PS—phenotypic selection; GS—genomic selection.

2. Methodology of Genomic Selection

GS is usually implemented in three or four steps: (1) genotyping and phenotyping of a “training population” from the breeding population of trees; (2) development of genomic prediction models where marker effects are simultaneously estimated for the alleles of all marker loci and which can predict phenotype from genotype; (3) validation of predictive models (on a “test population”, i.e., a group of individuals not included in model training); (4) application of the models to predict GEBV of non-phenotyped individuals and selection for further purposes based on these values [19]. Some authors, e.g., Tan [51], unite the first and second steps. Thus, a GS program uses two populations: a training population, for which genotypic and phenotypic data are determined, and a test population, where the breeding value of each individual is estimated based on the genotypic data.

Establishment of a training population is a critical step in GS [74]. A successful approach to establishing a training population in forest genomic breeding is to use available progeny tests from classical breeding programs. Such tests usually involve thousands of trees from several dozen (rarely hundreds) half-sib or full-sib families. Full-sibs produced from controlled pollination are preferable to open-pollinated half-sibs with only one known parent. However, half-sibs are widely used because they can offer some advantages: (1) fast and inexpensive production due to the lack of cross-breeding stage; (2) screening of a large number of parents with minimal efforts; (3) better genetic sampling due to a large number of pollen donors [61]. For minor tree species, genotypes from wild populations are also used. Training populations of forest trees usually include over a thousand individuals, i.e., much more than in GS of agricultural crops [11].

The plants of a training population are used for whole-genome genotyping through a large number (thousands to tens of thousands) of SNP markers and phenotyping for as many economically valuable traits as possible. The use of genome-wide markers is one of the fundamental features of GS. For each trait locus, there is a probability of being in LD with the marker locus, hence there is no need to look for specific significant QTL-marker associations, as in MAS [10]. In GS, phenotyping is only needed for building and validating genomic prediction models. Furthermore, with the same genomic profiles, the models can be easily recalibrated for other traits of interest in accordance with new breeding objectives [30].

Genomic prediction models are developed based on genotyping and phenotyping data. Such models are validated on a test population, which is only genotyped. The test population is usually related to the training population. Models with higher accuracy, i.e., correlation between observed and predicted breeding values, can be used to predict BVs in non-phenotyped individuals based on their genotypic data and ranking by this parameter. Plants with a high GEBV can be used for further breeding, and those with a high genomic estimated genotypic value (GEGV; where the prediction took account of non-additive genetic effects)—for clonal propagation and subsequent cultivation [24]. GS is focused on predicting general breeding value rather than identifying specific genes associated with a particular trait, and is therefore less limited by polygenic heritability of many traits in agricultural and forest species [12].

3. Genotyping Forest Trees

The development of high-throughput genotyping technologies has paved the way for a wide use of GS. A successful GS requires a dense and genome-wide distribution of genetic markers because in this case, each locus of a trait of interest is likely to be in LD with at least one marker locus in the entire target population [10]. In addition to the reduction in genotyping costs, the GS breakthrough was also facilitated by the development of statistical methods for accurate prediction of marker effects [75]. Efficient GS requires a low-cost, flexible, and accurate high-density genotyping platform [44]. GS of forest trees applies a number of genotyping technologies: SNP chip/array, diversity array technology (DArT and DArTseq), genotyping by sequencing (GBS), restriction site-associated DNA sequencing (RAD-seq), sequence capture, and genome-wide sequencing.

3.1. Single Nucleotide Polymorphic Marker Arrays

A single-nucleotide polymorphic marker (SNP) is a sequence with a single nucleotide replacement that does not change the overall length of the DNA sequence. SNPs have many applications in plants, including positional cloning, whole-genome association studies, mapping of QTLs, and determination of genetic relationships between individuals [40]. SNP array-based genotyping, with detection of the incorporated nucleotide on a two-dimensional surface (the chip), has become the most commonly used method in tree GS. Fixed SNP arrays provide the current gold standard for data reproducibility in forest tree GS; they are also breeder friendly, available from multiple service providers, and easily managed and stored [24].

Although two microarray-based genotyping platforms—Infinium (Illumina) and Axiom (Affymetrix)—were used for tree populations, most parts of GS research in trees were done using the Illumina Infinium platform. This genotyping platform for a non-model species (*Pinus taeda*) was first applied by Eckert et al. [76]. It contains 7216 SNPs, each representing a unique pine expressed sequence tag (EST) contig, and was used for GS of *Pinus taeda* populations [26,70]. For maritime pine, the Illumina Infinium SNP array with a large number of markers (9K and 12K) was used [7,18]. The genotyping of *Picea glauca* was done using two iSelect Infinium (Illumina) SNP arrays: PgAS1 [59] and PgLM3 [57]. Both had a similar number of assayed SNPs (13,162 and 14,139, respectively), but the first array was mainly designed for population genetics and genetic association studies, whereas the second one was constructed for population genetics, genomic prediction, and linkage mapping purposes [77]. SNP arrays were also developed for genotyping of minor species, e.g., the Infinium iSelect SNP genotyping array containing 5300 SNPs representing as many distinct black spruce gene contigs [64].

Based on the sequencing of 12 eucalyptus species, the high-density Illumina Infinium EuCHIP60K was developed, which contains probes for 60,904 SNPs [78]. According to the system's developers, EUChip60K represents an outstanding tool for GS, genome-wide association study (GWAS), and the broader study of complex trait variation in eucalypts. The chip was used in GS of various eucalyptus species and hybrids: *Eucalyptus globulus* [4], *E. grandis* × *E. urophylla* [42], *E. nitens* [47]. However, population screening for traits with oligogenic effects, such as disease resistance genes, can supposedly be done with a low-density SNP chip [1].

The Axiom system was used later and much less often: e.g., for genotyping *Cryptomeria japonica* [37] (more than 73,000 SNPs) and *Pinus contorta* [66] (more than 51,000 SNPs).

3.2. Diversity Array Technology

Diversity Arrays Technology (DArT) is yet another microchip-based marker system. This cost-effective sequence-independent ultra-high-throughput marker system was developed in 2001 [79]. The DArT technology is based on detection of polymorphic DNA fragments and, unlike SNPs, its chip development does not need genome sequence data. Hence, GS can also be used for species with an unsequenced or totally unstudied genome. This technology is also used for large-genome species such as grain crops [80]. Comparative evaluation of SNPs versus DArT markers in GS of wheat (*Triticum durum* Desf.) showed similar prediction accuracies of the two marker systems, although the number of DArTs was 3 times that of the SNPs (16,383 vs. 5649, respectively) [81]. Later on, there appeared an improved technology, DArTseq, which has largely replaced the original DArT. DArTseq is a DArT marker platform combined with next generation sequencing (NGS) platforms, which allows reduction in genome complexity mediated by restriction enzymes and sequencing of restriction fragments [82]. Comparative evaluation of various array-based platforms—DArT, Illumina Infinium BeadChip wheat iSelect, and DArTseq—in GS of spring wheat showed that DArTseq—given the low cost per SNP—was the best platform for genomic prediction [83].

SNPs and DArT markers were not compared on forest trees, but DArT was used on *Eucalyptus grandis* [41,43] and various eucalyptus species and hybrids [22], whereas DArTseq was used on *E. robusta* [49] and *E. urophylla* × *E. grandis* [50].

3.3. Complex Genome Genotyping

Coniferous species are characterized by large genomes of about 20 Gb or more (e.g., the *Pinus taeda* genome is 22 Gb [84]), which many times exceed the genomes of deciduous trees (e.g., 485 Mb for poplar [85]). The huge genomes of coniferous species impede their GS. Although the number of SNP markers in tree GS widely varies (from 2–3 to several dozens of thousands), most studies use no more than 10 thousand. This is quite enough for deciduous tree genomes, but not for conifers, where such a small number of markers may not be able to capture most of the QTL effects in large breeding populations [54]. To achieve the same coverage density as in deciduous species, one would need at least 100,000 to 200,000 markers. To save costs and time on conifer genotyping, Neves et al. [86] proposed focusing on the coding region. This method—exome capture—is a target enrichment method for sequencing the protein coding regions in a genome, which greatly improves the analysis efficiency. Genetic variation is not limited to exomes, but sequence capture is a cost-effective alternative to whole-genome sequencing. Sequence capture consists of targeted hybridization and allows genotyping almost any number and density of genetic markers. The method has become popular quite recently, after Suren et al. [87] showed the utility of sequence capture for re-sequencing in complex genomes of interior spruce (*Picea glauca* × *P. engelmannii*) and lodgepole pine (*Pinus contorta*). Sequence capture genotyping may be a useful alternative to DNA chips in GS of forest trees. It was used for genotyping *Picea abies* [54–56] and *Pinus radiata* [67], and was the only method applied in all GS studies on Douglas-fir [19,63,72]. A comparison of two genotyping methods—sequence capture followed by next generation sequencing (NGS) versus EucHIP60K.br—showed their equivalence in terms of genomic prediction of the traits of interest in eucalyptus [44]. The authors concluded that sequence capture could be a good alternative for species where SNP arrays are not available or too expensive to develop. This was later confirmed by Ballesta et al. [39], where the efficiency of EUChip60K on *Eucalyptus cladocalyx* was as low as 6% (about 3900 SNPs), probably due to a distant relatedness with the species on which the chip was developed.

An alternative to genotyping complex genomes may be provided by DNA sequencing associated with the restriction site (RAD-seq), which genotypes regions in proximity to endonuclease restriction sites [44]. This method allows a relatively cheap identification of a large number of SNPs required for GS in species with an unsequenced genome, and the number of loci selected depends on the number of these restriction sites. Fuentes-Utrillo [65] applied this method for GS of *Picea sitchensis*, with four restriction enzymes and the number of markers for mapping ranging from ~2000 to ~56,000, depending on the enzyme.

3.4. Next Generation Sequencing Technologies

Yet another method—genotyping-by-sequencing (GBS)—is based on advances in next generation sequencing technologies for obtaining SNP data and has very low per sample costs [80]. The GBS employs restriction enzymes to reduce genome complexity, and a barcoding system for multiplex sequencing. GBS does not require a decoded genome and is suitable for non-model species, such as forest trees [14]. Furthermore, GBS uses the genotyped population to detect markers, thus minimizing the ascertainment bias. A comparison of DArT versus GBS marker platforms in winter wheat showed that 38,412 GBS SNPs provided a higher GS accuracy than 1544 DArT markers. Noteworthy, the per sample cost of DArT genotyping was 2.5 times as high as that of GBS. The authors suggested that the higher accuracy of GBS was mainly due to the large increase in the number of markers. GBS was used in GS of both coniferous—*Picea engelmannii* × *P. glauca* [14,15], *Pinus sylvestris* [68]—and deciduous species—rubber tree [53], *Castanea dentata* [28].

Finally, the use of whole-genome sequencing (WGS) in GS is quite ambiguous: its high cost is not always compensated by a higher accuracy of GEBV due to increased marker density. This method was used for genotyping *Eucalyptus polybractea* containing a high concentration of desired oil composition in leaves [8]. Increased SNP density from 10K to 500K generally results in increased predictive ability for most traits tested, although for height, chip-based genotyping may be more cost-effective than

WGS. Apart from its use on eucalyptus, the same method has been recently employed to genotype *Shorea platyclados*, a tropical tree from the Dipterocarpaceae family with an excellent timber quality. This was the first use of GS for wood species from the Southeast Asian rainforest [73].

The use of high-density molecular markers can uncover hidden relatedness in open-pollinated populations and correct potential pedigree errors. For example, the average pairwise estimates of genetic relationship among individuals were substantially lower using SNP data than expectations based on pedigree information of eucalyptus hybrids [51]. The authors suggest that these inconsistencies likely derived from pollen contamination and/or mislabeling in the process of generating the full- and half-sib families. Pedigree errors are common in breeding programs and result in incorrect estimates of heritabilities and decreased breeding value accuracies. The accuracy of GEBVs was 0.55–0.75 when using the documented pedigree of the radiata pine training population and 0.61–0.80 when using the SNP-corrected pedigree [67]. The documented pedigree was corrected using a subset of 704 SNPs and about 50% of parents were reassigned. This pedigree error was significantly higher than that reported in livestock and plant breeding programs (about 10%) [67]. Errors were also found during the verification of the lodgepole pine's population structure [66]. Correct pedigree information is essential for accurate selection of suitable individuals as parents of the next generation.

The advent of NGS technologies has sharply reduced the cost of sequencing and allowed decoding the genomes of not only model species and important agricultural crops, but of some forest trees as well. With the exception of the genome of the model species *Populus trichocarpa* Torr. and A. Gray ex. Hook [85], all other sequenced tree genomes were published in the last decade. In total, the genomes of less than 10 coniferous and less than 20 deciduous forest tree species have been sequenced so far. They do not include Scots pine (*Pinus sylvestris*), an economically important species dominant in forested areas [88]. The availability of a sequenced genome is not a prerequisite for GS, although it would simplify the genotyping process. On the other hand, there are several important tree species with known genomes that have not yet been subject to GS (e.g., birch, oak), and they may probably be used to develop genotyping systems for forest breeding.

4. Phenotyping Forest Trees

4.1. Problems of Classical Phenotyping

The mechanisms of tree phenotype formation have their own specific features: during their long lives, trees are exposed to alternating dormancy/growth periods and lots of stresses including those never experienced by annual plants, e.g., autumn and late spring frosts. Poor understanding of phenotype formation mechanisms in forestry can lead to unpredictable behavior of some genotypes under certain environmental conditions [89]. Tree phenotyping should be done in different environments because the phenotype is influenced by both environmental and genetic factors.

Phenotyping has always been a bottleneck in classical breeding programs, since the size of breeding population is limited by the evaluable number of plants for which the phenotype can be determined. Genomic selection is no exception, and the phenotyping of the training population is as important as its genotyping. Precise phenotypic data are key to accurate GEBV prediction for the training population [90]. While plant genotyping once used to be the bottleneck, the development of NGS has reduced the costs of genotyping and now, the limiting factor is phenotyping. The current technical challenge for implementing GS in crop plants is the reliability of phenotypic data [20]. Traditionally, plant phenotypic traits were determined by time-consuming, labor-intensive, and often destructive manual measurements, which were also prone to researcher bias. The situation has changed with the recent advent of high-throughput phenotyping (HTP) methods that are an integral part of plant phenomics and are based on obtaining images in various spectral regions and their subsequent computer processing [91]. It is the phenomic approach that, for the first time, has provided an opportunity to deduce the patterns of changes in plant phenotypes in response to changing environmental factors. These methods allow non-invasive precise phenotyping of large

number of plants over long periods of time. Plants can be studied at a whole plant level (holistic phenotypes) or at an organ level, e.g., leaves and stems (component phenotypes) [92]. These methods assess various parameters automatically, thus eliminating the issues of researcher bias and inadequate statistical processing.

4.2. High-Throughput Phenotyping

A variety of equipment has been developed for HTP: individual devices, including robotic ones, as well as automated greenhouse systems that can quickly scan and record accurate data for thousands of plants by means of non-invasive image capture techniques in various spectral regions. Special equipment has been also developed for plant phenotyping in the field: moving platforms (“phenobiles”) and aerial systems (drones, unmanned aerial vehicles, balloons). Thus, the traditional manual techniques of plant phenotyping are now giving way to high-precision non-destructive imaging techniques. Furthermore, specialized image processing software has been developed to assess plant growth and development both at specific time points and throughout their life. Computer image analysis has significantly increased the phenotyping throughput that used to be the limiting factor in the analysis of genotypes or their behavior [93]. For instance, the Quantitative Plant database [94], formerly called Plant Image Analysis, contains nearly 200 plant image analysis tools.

HTP is most often based on visible-range images that provide much information about the plant’s height, the structure of its parts, and the size, color, and spatial orientation of its leaves. The use of infrared (IR) cameras to capture night images allows round-the-clock measurements. According to some studies, the spectral reflection of leaves closely correlates with their nitrogen and chlorophyll content, and it can be remotely measured with hyper- and multispectral devices, as well as a chlorophyll meter (SPAD meter) [95]. Therefore, it is possible to determine not only the rate of plant growth but also their physiological state, resistance to diseases, etc. Studies of the relationship between drought severity and associated physiological changes in plants showed the particular importance of long-term experiments [96]. Obtaining a true picture would require collecting phenotypic data at regular intervals throughout the plant’s life cycle.

For several years, HTP methods have been used in the GS of annual crops, grains in particular. Rutkoski [97] observed that aerial measurement of canopy temperature and vegetation index, as secondary traits, with a thermal and hyperspectral camera, increased the prediction accuracies for wheat grain yield by 56% in pedigree, and by 70% in genomic prediction models. Juliana et al. [98] used aerial HTP platforms in the wheat genomic selection for grain yield under stress conditions (drought, heat) and found that it increased selection intensity due to the large populations used. An unmanned aerial vehicle (UAV) was first used to assess the heritability of vegetation index traits measured on small unreplicated plots and to estimate the extent to which they are predictive of wheat grain yield in replicated yield trials [99]. In that study, aerial HTP provided a substantially better response to selection for grain yield than conventional visual selection. Not only remote phenotyping but also other methods were used in GS of wheat. For example, a robotic field phenotyping platform measured plant height [100]. In another study, grain yield was assessed from spectral reflectance data collected with a handheld multiple spectral radiometer [101]. HTP was also used in GS of other annual crops. Greenhouse-based HTP platforms were applied to measure shoot biomass and water use in GS of rice [102]. Processed images of cassava roots were used as a source of phenotypic data in genomic selection [103].

The developing technologies of precision phenotyping, remote sensing, robotics, and artificial intelligence enable breeders to perform high-throughput, low-cost, and labor-saving precision phenotyping. This can help scale up the experiments, reduce labor costs, and eliminate human errors in manual measurements [104]. However, HTP methods are still rarely used in GS of tree species.

4.3. High-Throughput Phenotyping in GS of Forest Trees

Large training populations (about 1000 trees) of woody plants make their phenotyping a long, complex, and expensive process. Apple breeders note that the development of gene technologies has brought about the big problem of HTP of large populations [1]. Thus, the main challenge in GS of trees is field phenotyping of large numbers of plants under various environmental conditions. The challenge cannot be met without HTP. Furthermore, tree breeding populations do not differ much from the wild ones, i.e., forest species are in early breeding stages. Meanwhile, selection is most effective at early breeding cycles, when the frequencies of favorable alleles in the target population are low, and here, the precision of phenotyping is critical. Otherwise, it will lower genetic gains because of the very low frequency of favorable alleles [105].

To the best of our knowledge, GS of forest tree species, in contrast to annual crops, hardly ever uses direct HTP techniques, nor are they often applied in forestry. UAV phenotyping was first used on a black poplar (*Populus nigra* L.) population of 503 genotypes on an area of 1.67 hectares, where two water treatments (well-watered and moderate drought) were compared [106]. The study showed that such a phenotyping technique can be an important tool for improving the efficiency of forest tree selection for climate change tolerance. Later on, UAV was used to estimate the total height, intra-annual height growth, and phenology of individual 15-year-old Norway spruce trees (*Picea abies*) in a dense field trial [107]. The method proved to be a cost- and time-effective alternative to manual height measurements, and its precision was high enough for selection with total tree height as the target trait. An automated phenotyping platform was successfully used to measure a variety of growth parameters and response to dry-down period in seedlings of two oak species (*Quercus bicolor* Willd. and *Q. prinoides* Willd.) under greenhouse conditions [108]. The use of image processing demonstrated changes in the leaf shape and size in some transgenic lines of aspen (*Populus tremula* L.) carrying the recombinant xyloglucanase gene in the second year of vegetation under semi-natural conditions [109]. Of particular interest are image-based methods that allow reconstructing of the 3D structures of the trunk and branches. Tree architecture is one of the main traits in GS of forest trees (Table 2), but its traditional assessment is very labor-intensive and imprecise.

There are hardly any known studies on direct HTP of forest species, whereas indirect phenotyping methods, such as near-infrared reflection spectroscopy (NIRS) and X-ray diffraction, are used in GS for high-throughput measurements of chemical and physical properties of wood. Although data obtained with such methods may not reflect the state of the whole tree, they are usually quite sufficient for GS [11]. For example, NIRS measurements of chest-level samples of *Acacia crassiparpa* A. Cunn. ex Benth. wood correlated well enough with those of samples taken along the entire tree height, so that tree breeders would still effectively select the best-ranked genotypes [110].

NIRS combines spectroscopic techniques and mathematical algorithms for indirect measurement of concentrations of OH-, NH-, CH-, or SH-containing compounds and is widely used to determine the content of nitrogen, moisture, carbohydrates, amino acids, and some other plant compounds in several crop species [90]. This method allows fast and precise assessment of several phenotypes at a time, thus saving wet chemistry costs [111]. In one of the first GS studies on trees, NIRS was used for indirect measurement of eucalyptus pulp yield [22], and later, it was also used to measure physical properties of wood, such as fiber length, coarseness, and number of fibers per gram [44], basic density [51], as well as chemical properties of wood, such as α -cellulose content, syringyl to guaiacyl lignin monomer ratio [46].

Density is one of the main characteristics of wood and is traditionally assessed by extracting samples of increment cores from trees and then, measuring their volume and weight. The method is relatively simple and accurate, but also time-consuming and labor-intensive [112]. An alternative method for measuring wood density is X-ray densitometry, which was often used to phenotype various spruce species in GS [58,59,64]. Another method of X-ray structure analysis, X-ray diffractometry, is used for microfibril angle (MFA) measurement [113]. Wood modulus of elasticity (MOE) can be calculated from wood density and MFA [58].

4.4. Importance of Age-Related Phenotyping

The age of phenotyping is very important for trees, which is not so for agricultural crops. According to published data, phenotyping age can vary in a wide range (Table 2): from 1–2 [8] to 40 years [14]. It is known that correlation between the same traits in young and mature trees can differ significantly depending on trait, species, environment, and age [114]. This was also observed in GS studies. Studies on *Pinus taeda* showed that models developed for growth traits based on data from 1–2-year-old plants had a limited accuracy in predicting phenotypes at 6 years [26]. Ratcliffe et al. [14] compared the heights of interior spruce trees at the ages of 3, 6, 10, 15, 30, and 40 years. The prediction accuracy of GS models based on 30-year height was nearly equivalent to those based on 40-year, but it significantly decreased with an increasing age difference between the training and test populations. On the other hand, quite opposite results have been obtained lately. In order to reduce the breeding cycle, Alves et al. [71] integrated indirect phenotypic selection based on greenhouse phenotypes with traditional GS. Plants of the same genotypes of *Populus deltoides* were grown in the field and in a greenhouse, with plant height measurements collected from week 1 to 15 in the greenhouse and at years 1 to 5 in the field. The study showed a moderate correlation (0.39–0.42) between greenhouse height measurements at weeks 13 and 15 and field measurements at years 3–5. By combining multiple greenhouse phenotypes into a selection index, a relative efficiency of ~0.48 was achieved, which is comparable with the results of GS models based on field phenotypes only [71].

The results of the study offer promising prospects. In GS of trees, phenotyping is almost always performed in the field. Exceptions are very rare and relate to specific traits, e.g., the rooting ability of *Pinus taeda* cuttings [69]. The phenotype is influenced by both the genotype and the environment and therefore, data obtained for naturally perennial plants grown under controlled conditions cannot be extrapolated to their behavior in the nature. However, the manifestation of some traits, such as tolerance to biotic and abiotic factors, are difficult to assess in the field, because it is difficult to provide a large variability of stress factors in order to evaluate the resistance of genotypes. Although pest and disease resistance can still be evaluated using artificial infection, practical, GS uses indirect methods. In particular, the levels of acetophenone aglycones (piceol and pungenol) that are known to be active ingredients against spruce budworm were analyzed in white spruce needles [31], and the growth traits in Norway spruce plants were assessed in areas with different pest abundance [30]. A trait such as disease resistance was evaluated under greenhouse conditions [34,69]. It is problematic to assess drought tolerance in plants, which, unlike some agricultural crops, are normally not artificially irrigated. Drought in field conditions is hardly predictable, and even more so is its duration and intensity. In a study on rubber tree [53], plants were cultivated on sites with different water availability conditions, but such sites would differ not only in the amount of precipitation, but also in temperature, soil composition, and many other factors. Meanwhile, drought tolerance is becoming an increasingly valuable trait for forest trees due to ongoing global climate change. The first study on GS for drought tolerance in trees was conducted not long ago on a eucalyptus hybrid population using water use efficiency (WUE) as a selective trait [32] and there is no doubt that research in this area will expand.

We think that GS for tolerance to abiotic stresses can use clonally propagated planting material from training populations to assess the effects of external factors on plant growth and development (in a greenhouse or, even better, outdoors). The level of correlation between the manifestations of such traits in juvenile and mature plants is unknown but can be roughly estimated from the growth history of the adult population and climate data. Traditional assessment of the physiological state of plants is long and labor-intensive. Bouvet et al. [32] assessed WUE in eucalyptus by measuring stable carbon isotope ratio ($\delta^{13}\text{C}$) in plant tissues by isotopic mass ratio spectrometry, which allows large-scale screening of plants. An alternative method is HTP using various spectral cameras. These techniques can effectively evaluate the growth, biophysical, and biochemical performance of plants [115] and will help reduce the cost of phenotyping and improve breeding value prediction accuracy for forest trees.

5. Genomic Prediction Models

5.1. Parametric and Nonparametric Models

GEBV estimation is no less important in GS than genotyping and phenotyping of individuals in a population. A genomic prediction method should provide a high accuracy and preferably capture LD between markers and QTLs rather than relatedness for higher long-term stability [11]. In addition, it must be a simple, reliable, and efficient tool for estimation of various traits. There are a number of statistical methods developed for use in GS. They mainly differ in the assumptions of the distribution and variances of marker effects. These models usually contain a huge amount of genotypic data (markers) and a limited amount of phenotypic data. All these methods can be divided into two groups: parametric and nonparametric models [10,111].

Parametric models were the first to be applied in GS. In the first GS study, Meuwissen et al. [16] used three such models: Ridge Regression Best Linear Unbiased Predictor (RR-BLUP), Bayes A, and Bayes B. RR-BLUP assumes that all marker effects are normally distributed and that all markers have equal variations with small but non-zero effect. The other two models are based on Bayesian estimation. In contrast to normal distribution in RR-BLUP, marker effects in the Bayes models are a priori assumed to follow Student's *t*-distribution [111]. Later on, modifications of RR-BLUP were developed (GBLUP (genomic BLUP), HBLUP (single-step genomic BLUP), as were a number of Bayesian models (Bayes $C\pi$, Bayes LASSO (least absolute shrinkage and selection operator), etc.).

GEBV prediction accuracy depends, among other things, on the statistical methods used. They differ in their assumptions and algorithms regarding the variance of complex traits with different genetic architectures. There are two types of genetic architecture: (1) genetic effects follow a mixed inheritance process where there are few genetic variants of large effects and many variants of very small effects, or (2) each genetic effect contributes only a very small fraction of the total genetic variance [67]. Bayesian models are better suited for traits, with the first type of genetic architecture, where marker effects are modeled to follow a prior distribution. The basic difference between various Bayesian methods is that they have different prior distributions and produce different degrees of shrinkage [116]. In particular, Bayes A assumes that genetic variance follows an inverted chi-square distribution and therefore, this model is suitable for traits controlled by a moderate number of genes; Bayes B assumes the variance of markers is equal to zero with probability π and is better suited when the trait is strongly influenced by certain loci; Bayes $C\pi$ assumes that probability π has a prior uniform distribution and therefore, it is better suited for analyzing real data [69,116]. The drawback of the Bayesian methods is the need to set prior values, but this requirement is circumvented in the Bayesian LASSO that requires less data [117].

For traits with the second type of genetic architecture, i.e., influenced by a large number of minor genes, better prediction accuracies are achieved with models like GBLUP and RR-BLUP, which assume that the effects of all loci have a common variance. GBLUP is equivalent to RR-BLUP and uses a genomic relationship matrix (GRM) generated by evaluating marker covariance across all individuals, providing greater resolution in genetic relationships among individuals [111]. This model can be applied even with simple or absent pedigrees and is therefore preferred for breeding programs of forest trees without a long record of commercial cultivation [67].

In one of the first GS studies, Bayesian approaches outperformed RR-BLUP for disease resistance traits controlled by a limited number of loci [69]. However, GS for disease resistance is very rare, and most studies of complex quantitative traits failed to find any advantages of certain models. For example, both RR-BLUP and Bayes $C\pi$ performed consistently well in a study on interior spruce [14]. A study on eucalyptus [4] showed comparable prediction accuracies of GBLUP and Bayesian models (Bayes B, Bayes C, BLASSO). The absence of marked differences between GBLUP and Bayesian methods in the evaluation of growth and wood quality traits was also observed in *Pinus pinaster* [18] and *Pinus contorta* [66]. In *Cryptomeria japonica*, however, GBLUP provided higher

prediction accuracies than Bayes B for almost all assessed traits, which suggests their control by many QTLs [37].

Unlike phenotypic mass selection based on an ancestral relationship matrix (matrix A), genomic prediction relies on a marker-based relationship matrix (matrix G), which provides a more accurate assessment of genetic similarity [56]. Generally, the studies showed the superiority of marker-based models over pedigree-based models. In a eucalyptus hybrid population [50], prediction accuracy and stability were improved by using marker-based instead of pedigree-based relationship matrices. In *Picea glauca*, marker-based models (GBLUP) showed better prediction accuracies compared to pedigree-based models (ABLUP) [59]. However, the advantage of such models was not always observed. In Norway spruce, ABLUP had higher accuracy for all four traits than four genomic selection methods (GBLUP, BRR, BLASSO, and RKHS (reproducing kernel Hilbert space) [54]. In a recent study on radiata pine [67], the accuracy of GEBVs (from a GBLUP model) was higher than that of EBVs (from an ABLUP model) for branch cluster frequency, but lower for stem straightness, internal checking, and external resin bleeding.

In addition to parametric models, GS also uses a group of nonparametric and semi-parametric models. Examples of such models are random forests (RF), support vector regression (SVR), and neural networks (NN), nonparametric machine learning methods, and RKHS, a semi-parametric method where the genomic relationship matrix used in GBLUP is replaced by a kernel matrix, which enables nonlinear regression in a higher-dimensional feature space [118]. Unlike parametric models that evaluate additive genetic effects, these models can also capture non-additive ones (e.g., dominance, epistasis) [111]. Thus, they can predict phenotypes better than the parametric models, especially where non-additive effects are important.

Nonparametric models are rarely used in forest tree GS, and only in recent years. In eucalyptus, RKHS demonstrated slightly better predictive abilities than four other models for traits with lower heritabilities (such as trunk CBH, height, and volume), but it was the worst for pulp yield [51]. In other studies, RKHS did not differ in accuracy from parametric statistical methods (GBLUP and BLASSO on Norway spruce [54] and BLASSO and RR-BLUP on rubber tree [27]). As far as we know, RF has only been used in studies on *Cryptomeria japonica* [13,37]. The prediction accuracy depended on where the plants were grown, but on the whole, GBLUP and RF were better models than Bayes B [37].

In each GS model, prediction is based on different analytical assumptions, hence there is no universally applicable statistical method for any traits in any population. Heslot et al. [119] compared 11 different parametric and nonparametric models for predicting various quantitative traits on wheat, barley, maize, and *Arabidopsis* and no model was better than the others for all traits. Normally, one should start with the use of RR-BLUP or GBLUP, include Bayesian models for large-effect loci, and machine learning methods for important non-additive effects [11].

5.2. Non-Additive Genetic Effects

Additive effects are considered the most important in breeding, since only they can be inherited, as opposed to non-additive effects associated with specific genotypes. However, dominance and epistasis can be confused with additive or random environmental effects, and if they are significant, ignoring them will lead to bias in genetic estimation. On the other hand, their inclusion in models can improve the accuracy of prediction [34]. Dominance assessment is important for crossed populations commonly used in the breeding of perennial species. Parametric models can also be used to estimate non-additive effects by replacing pedigree-based relationship matrices due to non-additive effects with their marker-based counterpart [70].

Early studies in forest tree GS evaluated only additive effects, but later simulation studies on eucalyptus [33] and loblolly pine [34] showed that the inclusion of dominance in the GS prediction model improved its accuracy. This was verified in practice by some studies. Bouvet et al. [50] showed that non-additive variance explained a significant part of the total genetic variance, although epistatic variance could not be clearly estimated. The authors attributed this to their use of

Eucalyptus grandis × *E. urophylla* hybrids that could have enhanced heterosis. In another study on eucalyptus, inclusion of dominance effects in the model increased the accuracy of GS for traits with a large dominance variance (height and CBH) [52]. The role of non-additive effects was also confirmed on coniferous trees: El-Dien et al. [59] reported a high proportion of epistatic variance in wood density of white spruce. Comparing GBLUP-A with GBLUP-AD on interior spruce showed the superiority of the latter model, which was more pronounced for height than for wood density due to the observed dominance variance [61]. In contrast to wood density, height assessments showed no differences between GBLUP-AD and GBLUP-ADE due to the lack of epistatic genetic variances. Thus, in interior spruce, tree height was best assessed with GBLUP-AD, whereas wood density with GBLUP-ADE, reflecting the presence of significant additive × additive genetic variances.

The role of non-additive effects was not always noted and often depended on trait. For instance, inclusion of dominance effects in BLR and RR-BLUP had no effect on GS accuracy in rubber tree [27]. Compared to an additive-only model, the use of an additive + dominance model in eucalyptus improved the predictive abilities for growth (mean annual increment) but not for wood quality traits [42]. Inclusion of dominance effects increased R² for tree height and acoustic velocity in Norway spruce both in the pedigree-based (ABLUP) and the genomic-based (GBLUP) models [55]. Yet, R² for Pilodyn penetration (a surrogate for the trait of wood density) and MOE did not change in either model, which was consistent with the zero estimates of dominance variations for both traits. Furthermore, the full genomic-based model with additive, dominance, and epistatic effects (GBLUP-ADE) almost did not differ from the GBLUP-AD model for all four traits (height, acoustic velocity, Pilodyn penetration, and MOE), which indicates the absence of three kinds of epistatic interactions in Norway spruce. Finally, as shown on eucalyptus hybrids, 0% to 30% of the phenotypic variance for growth traits could be attributed to epistatic variation depending on the plant age, and these findings are consistent with classical breeding studies in *Eucalyptus* [52]. However, epistasis showed no effect on basic density and pulp yield in eucalyptus, in contrast to El-Dien et al. [59], where the epistatic effect on wood density was noted. Thus, the contribution of non-additive effects may depend on plant species, trait, and age.

5.3. Multi-Trait and Multi-Environment GS

In classical breeding, tests under various environmental conditions are essential for assessing the influence of environmental factors on genotype behavior and for studying genotype by environment interactions ($G \times E$). $G \times E$ interactions are quite common in forest trees and are used to estimate the effectiveness of the same lines under different environmental conditions in genotype stability studies [11]. However, most GS studies use estimates obtained in only one environment, which prevents the use of information about environmental factors and thus, limits the predictive abilities of the GS model. Meanwhile, inclusion of genotype-by-environment interactions in GS models can improve their prediction accuracy. The effect of a genomic marker can be considered as a function of environmental covariates that can be evaluated by GS methods [120].

Genotype-by-environment interactions in trees are usually studied on contrasting sites [114]. Resende et al. [26] planted loblolly pine on four sites and found that the equation for any single site provided good prediction accuracy (0.64–0.74) within that site and a lower accuracy (0.18–0.66) for other sites. Site location was also important: where cross-validation was performed between sites within the same state (Florida or Georgia), the accuracy was higher compared with sites located in different states (0.54 and 0.37, on average). This finding was later confirmed on interior spruce planted on three sites: any of the two GS models, RR-BLUP and Generalized Ridge Regression (GRR), designed for a specific site predicted GEBV for other sites with a worse accuracy (0.41–0.59 vs. 0.00–0.39) [15]. Thus, $G \times E$ interaction was evident even though all three sites were located within one breeding zone. The researchers also used the multi-site population for prediction of the GEBV for each individual site. Although the accuracies varied by site, they were higher (0.42–0.49) than in the cross-site validation study [15]. A GS study on Norway spruce planted on two sites in northern Sweden showed that for all four traits, the accuracies of within-site training and selection were always higher than those of

cross-site training and selection [54]. A recent study on rubber tree demonstrated the superiority of multi-environment GS models over single-environment ones [53].

The effect of G×E may also depend on the trait. As known from classical breeding, G×E usually has a significant effect on growth traits in coniferous species, while having no influence on their wood quality [54]. The GS studies on white spruce planted on two sites in different ecoclimatic regions of Canada confirmed that growth traits were more sensitive to G×E than wood traits [57]. Similar findings were reported for Norway spruce [54]. Thus, a genomic model developed for one site can be used to predict GEBV for wood traits in another site, but one should bear in mind that G×E can have a stronger impact in a more heterogeneous environment. In general, the accuracy of a predictive model decreases when applied to a different environment, but the degree of the decrease depends on the trait. GS should also take G×E into account in order to select genotypes adaptable to changes in environmental conditions.

In addition to multi-environmental trials, multi-trait selection is also of considerable interest in GS. GS models are usually designed to predict only one trait. However, multi-trait models were also developed and helped improve the accuracy of BV predictions in animals [43], crops [116], and forest trees. Studies on *Eucalyptus grandis* demonstrated the superiority of the multiple-trait combined approach in predicting BVs over the single-trait combined approach, in particular for a low-heritability trait (height) [43]. As shown on a eucalyptus hybrid, positive genetic gains can be achieved by associating biomass, a proxy of WUE, and wood chemical traits (cellulose and lignin) [32]. Interesting results were obtained on interior spruce with multi-trait GS prediction models based on Principal Component Analysis (PCA) [15]. There is a known negative genetic correlation between wood yield and wood quality, which was confirmed by the results from PC1. However, it turned out that the use of PC2 and PC3 allowed a concurrent selection of traits with different phenotypic optima, i.e., PC2 and PC3 accessed different combinations of SNPs (i.e., causal genes) that work in the same direction [15]. All these studies show the feasibility of GS for several traits at a time, even with a negative correlation between them.

5.4. Epigenetic Effects

Along with genetic effects, there are also epigenetic effects, and there is growing evidence that epigenetics has the potential to contribute to important traits in many plant species. Epigenetic trait is a heritable change in gene expression without changes in DNA sequence. There are several epigenetic mechanisms, and the best studied one is the methylation of cytosines within CG dinucleotides [121]. These changes can occur much faster than the genetic ones and they respond to external stresses; therefore, they may be particularly important in the context of a rapid climate change [122]. The role of epigenetic changes in phenotypic plasticity has been increasingly studied in recent years. Studies of DNA methylation in response to climate change are of special importance for long-living trees because their long generation period limits their ability to respond to rapid environmental changes through genetic mechanisms [123]. For example, studies revealed an association between DNA methylation and climatic conditions in natural populations of valley oak (*Quercus lobata* Nee) [123] and a correlation of DNA methylation levels and biomass productivity of poplar plants grown under different water availability conditions [124]. In traditional plant breeding, there is such a phenomenon as transgressive segregation: the hybrid progeny has a wider phenotypic variation than its parents. Unlike heterotic phenotypes, extreme phenotypes caused by transgressive segregation are heritably stable, but the precise molecular mechanisms of this phenomenon remain unclear [125]. It is assumed that transgressive segregants in plant breeding that are likely due to genetic and/or epigenetic effects are more common for adaptive traits such as tolerance to abiotic stresses, and water and nutrient use efficiency that strengthen the plant viability [126].

Although epigenetic effects are considered to be inheritable (additive), they can also be non-additive. Such a phenomenon as heterosis, where the progeny (F1 hybrids) surpasses its parents in a number of traits, is often used in plant breeding, including forest breeding. However, despite the wide practical use

of heterosis, its underlying biological mechanisms are not well understood [127]. Epigenetic changes were shown to be involved in the generation of heterotic phenotype in annual plants [128]. Later, Gao et al. [129] investigated the role of DNA methylation changes in F1 hybrids of *Populus deltoides* and found non-additive levels of methylation, suggesting a role of DNA methylation in heterosis of forest trees. Thus, the stability of epigenetic changes (their heritability or non-heritability) may be of great importance in GS, as it will affect the degree of LD of epialleles with nearby SNPs [130].

Statistical discrimination between genetic and epigenetic changes underlying phenotype changes has not been studied well enough yet. The existence of epigenetic inheritance means that matrix A , which is used to describe inheritance in the calculation of estimated breeding values (EBVs), does not exactly describe the genetic similarity between relatives [131]. A recent application of Bayesian statistical models to assess the epigenetic architecture of complex traits in clinical practice showed that BayesRR distinguish between the variance explained by genetic markers and methylation probes better than LASSO or ridge regression [132].

It is already known that epigenetics can influence all aspects of the phenotypic variance in plants, including genetic variance, environmental variance, $G \times E$, etc. [121]. Clinical studies showed that DNA methylation profiles have the potential to significantly improve complex trait prediction over and above that of SNP markers [133]. Recent advances in genomic technologies and bioinformatics have made genome-wide high-resolution methylome analysis feasible even for very large genomes [134]. Thus, the use of these technologies in the GS of trees to detect epigenetic polymorphisms (markers) helps understand the role of epigenetics in phenotypic variation and improve predictive abilities.

6. Accuracy Drivers in Genomic Predictions

The efficiency of GS in plants was initially evaluated using computer simulations. The first study of this kind was performed by Bernardo et al. [135] in maize, and was soon followed by a simulation study on a woody plant, oil palm [136]. According to the latter, response to GS was higher than to MAS for all population sizes, and higher than to phenotypic selection for a population of 50 or 70 individuals. Simulation studies on forest trees [23,25] showed that GS was more efficient than traditional breeding. According to these simulations and the subsequent empirical studies, the accuracy of GS in forest trees depended on several key factors [19,51]: (1) the extent of LD between markers and QTLs, (2) effective population size (N_e), (3) marker density, (4) training population size, (5) relatedness between training and test populations, (6) genetic architecture of a trait (number of loci and effect size), (7) trait heritability. Of these, only the inheritance and the genetic architecture of traits cannot be controlled by the breeder, while the others can be, more or less.

6.1. Linkage Disequilibrium, Effective Population Size, and Marker Density

The first three of the above-mentioned factors are interrelated. Linkage disequilibrium (LD) is particularly important in the context of GS in forest trees [23]. GS uses the LD between markers and QTLs by employing a large number of DNA markers: with dense coverage, any QTL associated with a trait would be in LD with at least one marker [18,137]. Thus, GS prediction accuracy depends on the degree of LD, which, in turn, depends on the effective population size (N_e) of training population and the marker density.

Most tree genomes have a low level of LD, which is due to high outcrossing rates, long-distance propagule dispersal, and large effective population sizes [138]. Grattapaglia et al. [23] suggested that the low LD levels in most forest tree species could be increased by reducing the effective population size and increasing the marker density. N_e is one of the most important parameters in population genetics because it determines the effectiveness of natural selection [139]. N_e is also important in GS: the smaller N_e , the higher the predictive ability, since a lower N_e is associated with a higher LD, and hence, a stronger association between markers and QTLs over long distances throughout the genome [66]. It is believed that the success of GS in trees mostly depends on the interplay of two

factors: (1) N_e that determines the marker density and (2) LD that can be controlled by selecting N_e (it can be increased by reducing N_e) [140].

A simulation study on forest trees revealed a non-proportional inverse relationship between marker density and N_e : an adequate GS accuracy (≥ 0.68) was achieved with a marker density of about 2 markers per cM for $N_e = 30$, about 10 markers per cM with $N_e = 60$, and up to 20 markers per cM with $N_e = 100$ [23]. Similar results were obtained in a simulation in conifers (*Cryptomeria japonica*): efficient GS was achieved with one marker per cM [25]. Forest tree genomes are about 1000–2000 cM long: from 919 to 1814 cM for *Eucalyptus* species [141], 1637 cM for *Pinus taeda* [142], 1859 cM for Douglas-fir [143]. Thus, with a small N_e , the GS of an average genome would require as little as 3000–5000 markers, but even with a 2–3-fold larger N_e it would require several dozens or hundreds of times more markers, depending on the genome size.

Practically, non-domesticated forest trees are characterized by a large genetic diversity and a large N_e . N_e can be reduced by using populations of related individuals (half-sibs or full-sibs), as done in some tree breeding programs. Meuwissen et al. [144] found that adequate accuracy of predicted breeding values in populations of unrelated individuals can be achieved with a minimum of $10 \times N_e \times L$, where L is the total genome length in Morgans. Within-family GS requires much less markers, as they would track only the large chromosome segments shared by family members. For example, GS in a biparental population of apple trees showed a good accuracy (>0.7) with as little as 2500 markers [145], whereas theoretically it would have required 130,000 ($10 \times 1000 \times 13$) markers in an outbreeding population of apples [111]. A study on eucalyptus showed that a reduction in N_e from 51 to 11 improved the accuracy from 0.65 to 0.80 [22]. On the other hand, a large effective population size may lead to greater recombination and genetic diversity within the population due to high outcrossing rates in wind-pollinated species, which is favorable for a long-term genetic gain [35].

Experimental studies in forest tree species demonstrate quite fair predictive abilities at relatively moderate genotyping densities (2500–10,000 SNPs), probably due to the impact of relatedness as a driver of accuracy [11]. For example, in a study on *Picea abies*, the accuracy reached a plateau at 4000–8000 SNPs [55]. A subset of 3000–4000 markers was sufficient to reach the same predictive abilities and accuracies as the full set of 8719 markers in Scots pine [68]. If possible, however, higher marker densities should be used to increase the prediction accuracy. Furthermore, higher marker densities may be necessary where the training and test populations do not originate genetically from the same primary population [144].

6.2. Size and Structure of Tree Populations in GS

A larger training population allows more accurate assessment of marker effects and, hence, a higher accuracy of GEBV for candidate selection. Unlike some agricultural crops and even more so animals, for most forest trees, the size of training population is usually not a limiting factor. A simulation study showed that with a population size of more than 2000 individuals, the prediction accuracy leveled off to a plateau regardless of N_e and marker density [23]. However, with a high marker density or $N_e < 30$ individuals, 1000 individuals were enough to achieve accuracies of at least 0.7. Therefore, the training populations included 800–1200 trees in most studies (Table 2). A training population is usually sampled from an existing progeny trial derived from interbreeding (open or controlled pollinated) a few dozen elite parents. The costs of genotyping have dropped in recent years and it is no longer the limiting factor in the analysis of large populations. Moreover, genotyping of a specific individual has to be done only once. On the other hand, the costs of phenotyping remain high because it requires ample manual labor and should preferably be repeated in different years, environments, and in several replications. Therefore, it is phenotyping that is becoming the limiting factor for increasing the size of training population. The use of HTP methods will help remove this limitation.

In theory, a small N_e (related individuals) should provide higher prediction accuracy than unrelated genotypes. Calculations show that the use of 1000 eucalyptus half-sibs potentially has a very high predicted GEBV accuracy (>0.9) owing to the relatively small effective number of independent

chromosome segments within the half-sib families [111]. To test this statement, the effectiveness of GS was evaluated in a white spruce (*Picea glauca*) population of a large effective size [58]. The study showed that the accuracy of GEBVs obtained with individuals of unknown relatedness was lower, with about half of the accuracy achieved with half-sibs.

GS studies on forest trees generally reported moderate to high accuracies of selection models with correlations from 0.6 to 0.8 for full-sib-families, from 0.3 to 0.5 in half-sibs, and a very low predictive ability for unrelated individuals [64]. Furthermore, the full-sib family structure required significantly fewer markers than did the half-sib family structure to achieve the same effect. For example, in the full-sib family of Norway spruce, 250 markers provided the same accuracy, and 750 markers provided the same predictive ability for all four traits, as 100,000 markers in the half-sib family structure [54].

The degree of relatedness in a training population affects the prediction accuracy and is an important factor in the relationships between training and test populations, as was demonstrated in a number of tree studies. In these studies, the progeny of full- and/or half-sibs within one generation was usually divided into training and test populations [56]. This improved the accuracy of GEBV prediction by increasing the likelihood of the same chromosome segments being present in both populations [111]. On the contrary, the use of unrelated populations reduced the accuracy of predictive models. For example, models built with a half-sib structure of *Picea mariana* (Mill.) Britton, Sterns, and Poggenb. led to a large decrease in accuracy, predictability, and genetic gain compared with a full-sib design [64]. In a study on eucalyptus, the average values of the realized genomic relationships among full-sibs, half-sibs, and unrelated individuals consistently decreased (0.309, 0.131, and 0.0056, respectively).

In addition to the degree of relatedness, the ratio of training to test population was also important. In a study on Norway spruce, Chen et al. [54] compared the effects of five different training to test ratios (1:1, 3:1, 5:1, 7:1, and 9:1) on the accuracy of statistical models. The GS accuracy increased with the increasing ratio, although it also depended on the evaluated trait: for tree height, the maximum accuracy was achieved at the ratio of 5:1, whereas for wood quality traits, only minor improvements were observed after the ratio had been increased to 3:1. In a similar study on a deciduous species, eucalyptus genotypes were divided into five different size groups with the training to test ratios of 1:1, 2:1, 3:1, 4:1, or 9:1 [51]. In contrast to the findings of Chen et al. [54], here, the predictive ability significantly improved after the ratio had been increased from 1:1 to 9:1 (which supports the importance of an adequate size of the training set in GS), but it did not depend on trait. However, the ratio increase from 1:1 to 2:1 caused a greater increase in predictive ability than the ratio increase from 2:1 to 9:1. Thus, it may be more appropriate to improve the predictive accuracy by increasing the size of training population rather than the marker density.

6.3. Heritability and Genetic Architecture of Traits

Trait heritability and genetic architecture also affect model accuracy, but, unlike the above key factors, they are natural features and cannot be controlled by the breeder. To some extent, they can be influenced by choosing an adequate statistical model. Traits with a simpler architecture (e.g., disease resistance) have fewer loci that control large proportions of phenotypic variance. These are the features that are best suited for MAS applications and are better predictable in GS. More complex quantitative traits (growth, wood properties) may be controlled by dozens or hundreds of QTLs with weaker effects [11]. Given these differences in marker effects, calculations can be done using models that provide for different or equal contributions of all markers to the observed variability. For instance, the growth and wood attributes of interior spruce were more accurately predicted by RR-BLUP than by GRR, thus suggesting the complex genetic architecture of the traits [15]. De Almeida Filho et al. [34] compared predictions of polygenic (height, 1000 QTLs) and oligogenic (disease resistance, 30 QTLs) traits in a simulated population of loblolly pine (*Pinus taeda*). Computations showed that the models Bayes A and Bayes B were more accurate than BL and BRR for oligogenic traits in all scenarios. Thus, RR-BLUP (frequentist version of BRR) is better suited for predicting complex traits than simple ones.

The heritability of a trait can be defined as the fraction of phenotype variability which is due to genetic variation. Narrow-sense heritability considers only additive genetic effects, while disregarding the non-additive effects (dominance, epistasis) and genotype-environment interaction [111]. Traits with a higher heritability are more accurately described by marker effects and are less dependent on other factors. Heritability was shown to have a relatively small impact on the accuracy where the training population was large enough for adequate assessment of marker effects [11]. However, in the same way as for genetic architecture, modeling of dominance effects can improve the GS models for some traits [66]. The heritability of economically valuable traits is usually high enough to assure accurate GS of plants, provided there are sufficient markers and large training populations [111].

7. Economic Efficiency of GS in Tree Breeding

The current common breeding approaches—traditional, MAS, and GS—differ in cost and efficiency. Breeding of new varieties should consider the economic aspect in the process of cost-benefit analysis. Traditional breeding is based on phenotypic selection and has a long cycle and a low efficiency when selecting for complex traits [17]. Owing to early evaluation of the genotype, MAS can shorten the breeding cycle, which could otherwise reach 25 or more years in boreal conifers [146], but it is still ineffective for selection of polygenic traits. GS can shorten the breeding process even further, primarily by reducing long and expensive field trials, but this technology is less than 10 years old and its use is currently limited due to poorly studied genomes of many tree species. Nearly all studies that compared GS and traditional breeding of forest trees showed the superiority of GS, which significantly reduced the breeding cycle (Table 2). This was mainly achieved due to the possibility to evaluate plants at a very young age (down to seedlings). In loblolly pine, the efficiency of GS was 53–112% higher than with traditional breeding, and the breeding cycle shortened by 50% [26]. In radiata pine, the efficiency of GS was 37–115% higher, and the breeding cycle shortened from 17 to 9 years [67] compared with traditional breeding. The results in interior spruce were more modest than in pine species: efficiency increased by 6–33% and the breeding cycle reduced by 25% [14]. A twofold reduction in the breeding cycle was achieved in eucalyptus [22,46] and even threefold reduction in some *Picea* species [58,64] and rubber tree [53]. At the same time, GS relies on expensive genotypic analysis, and this must be taken into account when comparing breeding methods in terms of efficiency.

The economic efficiency of a breeding method may depend on such factors as population size, trait heritability, costs of plant phenotyping and genotyping, etc. For example, a simulation study on herbaceous plants showed that GS was more cost-effective than phenotypic selection if the following conditions were met: the heritability of traits of interest was below 0.25, and the cost of one plant phenotyping did not exceed that of genotyping [147]. Forest trees species differ in genome organization, duration of juvenile period, vegetative propagation abilities, and different rotation periods and requirements for growing conditions. The traits of interest for selection are also very diverse: quantitative (growth rate), qualitative (wood properties), and those responsible for tolerance to biotic and abiotic stresses. All the above can influence genotyping and phenotyping, propagation and cultivation during the breeding process, and makes each GS program unique.

The cost-effectiveness of GS in forestry depends on many factors and can vary greatly in each particular case. This technology is, perhaps, hardly suitable for small breeding programs with minor tree species and a poorly studied genome, but when used on major industrial forest species, its considerable time gain may be very important from a commercial perspective [11]. Resende et al. [22] calculated that the costs of GS of 20,000 eucalyptus seedlings would be paid back at least 20 times due to a 1% increase in pulp yield, and 9 years earlier than in the case of conventional breeding. On the other hand, genetic technologies are rapidly developing, and in the coming years, genome exploration may stop being a limiting factor.

The breeding of trees, even using GS methods, is a long and expensive process, and therefore, a detailed cost-benefit analysis is absolutely necessary before its practical implementation [11]. The very first GS study on woody plants (oil palm) [136] estimated that the cost per unit gain with GS was

26–57% lower than with phenotypic selection when markers cost USD 1.50 per data point, and 35–65% lower when markers cost USD 0.15 per data point. Only recently, however, there appeared full-scale studies on forest trees, with the assessment of various scenarios of breeding, propagation, plantation establishment, forest management, and harvesting. Chang et al. [148] conducted a stand-level financial benefit-cost analysis to compare GS and traditional breeding of two major commercial tree species in Western Canada, white spruce and lodgepole pine, if grown for up to 250 years. According to the results, GS could shorten the breeding cycle for these species in Canada from 33 to 18 years, but under current market conditions, traditional breeding should still remain the main tree breeding strategy for producing improved seedlings of white spruce and lodgepole pine for reforestation. Yet, GS would become a promising approach in the following scenarios: (1) an increased log price premium at harvest; (2) reduced seedling costs; (3) achieving higher genetic gain; and (4) planting on high-productivity sites. This study also shows the importance of choosing an appropriate reference year for comparing different breeding strategies, as this can significantly influence the conclusions made [148]. The authors emphasize that their calculations were made for the current market conditions and breeding strategies for white spruce and lodgepole pine in the province of Alberta, Canada, and cannot be extrapolated to other species and regions of the world. For example, a shorter rotation period and a lower cost of pine seedlings in the southern USA, New Zealand, or Chile may make GS financially more attractive.

Another study compared the financial performance of various breeding and deployment scenarios, with or without GS, in the context of intensively managed plantations of white spruce in Quebec (Canada) [149]. The duration of a classical breeding cycle, 34 years, was similar to the previous study, but the rotation period lasted for up to 60 years, and weeds were controlled with herbicides. According to the results of the study, the best scenario used GS with somatic embryogenesis (SE), followed by a scenario with GS in combination with top-grafting for the production of improved seedlings. As was already reported earlier, the most significant breeding cycle reduction with maximum increase in genetic gain could be achieved by combining GS and SE [150]. Li et al. [35] reported that the use of SE for propagation, even after traditional breeding, led to an additional 8.11% genetic gain compared with propagation by seeds of the selected individuals. On the other hand, combining GS with top-grafting resulted in a significant increase in additional genetic gain per year in conifers due to the reduced age of coning from 5 to 3 years [35]. Thus, top-grafting could be an interesting alternative to increase genetic gain and shorten breeding cycles for tree breeding programs where SE at an operational scale is yet not available [149]. Finally, the continuing reduction in genotyping costs is making GS increasingly financially attractive for use on forest trees.

8. Perspectives of GS in Forestry

Studies of GS in forest trees have been conducted for less than 10 years, but they have already demonstrated the promising prospects of this approach in forest tree improvement. A few years ago, Isik et al. [1] named three challenges for GS in trees: (1) lack of reliable, repeatable, and cost-efficient genotyping platforms; (2) lack of infrastructure to store, retrieve, and analyze large numbers of markers; (3) lack of well-designed and well-tested breeding populations for many species. The latest advances in genomics and bioinformatics can help solve the first two problems rather quickly, but the establishment of breeding population takes much longer. Nevertheless, the list of tree genera used in GS studies has significantly expanded over the last 2–3 years and is likely to continue expanding in the future. Among boreal species, good candidates for inclusion in the list are representatives of the genera *Betula* and *Quercus*. Russian breeders have been working with such species as *B. pendula* Roth, *B. pubescens* Ehrh., and *Q. robur* L. for several decades, and the breeding populations, including full-sib families, can be used for GS. Genotyping of these species is facilitated by the availability of the recently sequenced reference genomes of birch (*B. pendula*) [151] and oak (*Q. robur*) [152]. Their small genomes, 440 and 740 Mb, respectively, are ten times smaller than those of pine and spruce (~20 Gb), which is also an advantage. Birch and oak have valuable wood and they are very promising for plantation forestry in southern regions of European Russia. Of tropical species, *Tectona grandis* L.f., for which a draft genome

was recently published [153], can be considered as the best candidates for GS. It was already shown that the breeding cycle in coniferous trees can be further shortened by using grafting and somatic embryogenesis. In deciduous species, flowering can be accelerated with the fast-track breeding system using genetic transformation by flowering inducing genes. The system was successfully applied on fruit trees—apple [154], plum [155], and trifoliate orange [156]—and made it possible to reduce the juvenile period several times.

The list of traits of interest for GS will also expand. Global climate change is drastically changing the cultivation conditions of forest trees, especially in the boreal regions. On the one hand, increasing temperature can improve forest productivity. On the other hand, the duration and intensity of various stresses also increase, and traditional breeding fails to respond timely to these challenges. GS programs would probably also include such traits as tolerance to abiotic stresses, primarily drought. It is difficult and time-consuming to assess the state of plants exposed to stress factors using traditional approaches, hence the need for wider adoption of HTP techniques in GS, because they allow such assessment to be made remotely by capturing images in various spectral regions. Stresses are environmental factors that occur unpredictably and usually continue for a limited period of time, and can affect economically important traits of trees. Considering the above-said factors, GS programs should make a wider use of multi-environment, multi-age, and multi-trait statistical models.

In addition, the proper use of epigenetic variance can open up new opportunities for improvement of forest trees. This would require a detailed understanding of how to predict the stability of epigenetic variants (their additive or non-additive nature) so that epigenetics could be used to improve valuable heritably stable traits, primarily stress tolerance. In particular, changes in methylation of transposable elements can be responsible for the variability of traits [157]. This is especially important for conifers, whose large genomes contain a high number of transposable elements, which provide vast potential genetic resources [158].

Finally, we can expect more extensive use of GS simulations. The GS of forest trees began with simulation studies and such studies are still occasionally conducted. As shown by the recent cost–benefit studies, the economic efficiency of GS implementation depends on a large number of factors, including those difficult to predict (e.g., the market value of wood). GS programs are expensive and lengthy, although much shorter than traditional breeding, and their implementation should be preceded by simulation studies to select optimal strategies for specific species, traits, and growing conditions.

9. Conclusions

Despite its short history, GS has already proved to be a powerful tool in plant breeding. It is especially useful for the prediction of complex quantitative traits, including productivity and wood quality, the main characteristics of woody plants for cultivation on forest plantations. In each specific case, it is essential to consider a number of factors and their combinations with the biggest impact on the selection efficacy. Clarification of the pedigrees of individuals from breeding populations using molecular markers will improve the accuracy of predicting their breeding values. In general, studies on GS of forest trees confirm that it can significantly reduce the duration of the breeding cycle and increase the genetic gain. The reduction in genotyping and phenotyping costs is likely to contribute to a wider use of this method in forest breeding.

Author Contributions: Conceptualization, V.G.L.; writing—original draft preparation, V.G.L. and T.N.L.; writing—review and editing, V.G.L. and K.A.S.; funding acquisition, A.I.C. and K.A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was carried out within the state program of Ministry of Science and High Education of the Russian Federation (theme “Plant molecular biology and biotechnology: their cultivation, pathogen and stress protection (BIBCH)” (No 0101-2019-0037).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Isik, F.; Kumarx, S.; Martínez-García, P.J.; Iwataj, H.; Yamamoto, T. Acceleration of forest and fruit tree domestication by genomic selection. *Adv. Bot. Res.* **2015**, *74*, 93–124. [[CrossRef](#)]
2. Isik, F. Genomic selection in forest tree breeding: The concept and an outlook to the future. *New For.* **2014**, *45*, 379–401. [[CrossRef](#)]
3. Harfouche, A.; Meilan, R.; Kirst, M.; Morgante, M.; Boerjan, W.; Sabatti, M.; Mugnozza, G.S. Accelerating the domestication of forest trees in a changing world. *Trends Plant Sci.* **2012**, *17*, 64–72. [[CrossRef](#)] [[PubMed](#)]
4. Durán, R.; Isik, F.; Zapata-Valenzuela, J.; Balocchi, C.; Sofía Valenzuela, S. Genomic predictions of breeding values in a cloned *Eucalyptus globulus* population in Chile. *Tree Genet. Genomes* **2017**, *13*, 74. [[CrossRef](#)]
5. Muranty, H.; Jorge, V.; Bastien, C.; Lepoittevin, C.; Bouffier, L.; Sanchez, L. Potential for marker-assisted selection for forest tree breeding: Lessons from 20 years of MAS in crops. *Tree Genet. Genomes* **2014**, *10*, 1491–1510. [[CrossRef](#)]
6. Ballesta, P.; Serra, N.; Guerra, F.P.; Hasbún, R.; Mora, F. Genomic prediction of growth and stem quality traits in *Eucalyptus globulus* Labill. at its southernmost distribution limit in Chile. *Forests* **2018**, *9*, 779. [[CrossRef](#)]
7. Bartholomé, J.; van Heerwaarden, J.; Isik, F.; Boury, C.; Vidal, M.; Plomion, C.; Bouffier, L. Performance of genomic prediction within and across generations in maritime pine. *BMC Genom.* **2016**, *17*, 604. [[CrossRef](#)]
8. Kainer, D.; Stone, E.A.; Padovan, A.; Foley, W.J.; Külheim, C. Accuracy of genomic prediction for foliar terpene traits in *Eucalyptus polybractea*. *G3 Genes Genomes Genet.* **2018**, *8*, 2573. [[CrossRef](#)]
9. Zapata-Valenzuela, J.; Whetten, R.W.; Neale, D.; McKeand, S.; Isik, F. Genomic estimated breeding values using genomic relationship matrices in a cloned population of loblolly pine. *G3 Genes Genomes Genet.* **2013**, *3*, 909–916. [[CrossRef](#)]
10. Desta, Z.A.; Ortiz, R. Genomic selection: Genome-wide prediction in plant improvement. *Trends Plant Sci.* **2014**, *19*, 592–601. [[CrossRef](#)]
11. Grattapaglia, D. Status and perspectives of genomic selection in forest tree breeding. In *Genomic Selection for Crop Improvement*; Varshney, R.K., Roorkiwal, L., Sorrells, M.E., Eds.; Springer International Publishing AG: Cham, Switzerland, 2017; pp. 199–249. [[CrossRef](#)]
12. Zapata-Valenzuela, J.; Isik, F.; Maltecca, C.; Wegrzyn, J.; Neale, D.; McKeand, S.; Whetten, R. SNP markers trace familial linkages in a cloned population of *Pinus taeda*—Prospects for genomic selection. *Tree Genet. Genomes* **2012**, *8*, 1307–1318. [[CrossRef](#)]
13. Nagano, S.; Hirao, T.; Takashima, Y.; Matsushita, M.; Mishima, K.; Takahashi, M.; Iki, T.; Ishiguri, F.; Hiraoka, Y. SNP genotyping with target amplicon sequencing using a multiplexed primer panel and its application to genomic prediction in Japanese cedar, *Cryptomeria japonica* (L.f.) D.Don. *Forests* **2020**, *11*, 898. [[CrossRef](#)]
14. Ratcliffe, B.; El-Dien, O.G.; Klápště, J.; Porth, I.; Chen, C.; Jaquish, B.; El-Kassaby, Y.A. A comparison of genomic selection models across time in interior spruce (*Picea engelmannii* × *glauca*) using unordered SNP imputation methods. *Heredity* **2015**, *115*, 547–555. [[CrossRef](#)] [[PubMed](#)]
15. El-Dien, O.G.; Ratcliffe, B.; Klápště, J.; Chen, C.; Porth, I.; El-Kassaby, Y.A. Prediction accuracies for growth and wood attributes of interior spruce in space using genotyping-by-sequencing. *BMC Genom.* **2015**, *16*, 370. [[CrossRef](#)]
16. Meuwissen, T.H.E.; Hayes, B.J.; Goddard, M.E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **2001**, *157*, 1819–1829.
17. Heffner, E.L.; Lorenz, A.J.; Jannink, J.-L.; Sorrells, M.E. Plant breeding with genomic selection: Gain per unit time and cost. *Crop Sci.* **2010**, *50*, 1681–1690. [[CrossRef](#)]
18. Isik, F.; Bartholomé, J.; Farjat, A.; Chancerel, E.; Raffina, A.; Sanchez, L.; Plomion, C.; Bouffiera, L. Genomic selection in maritime pine. *Plant Sci.* **2016**, *242*, 108–119. [[CrossRef](#)]
19. Thistlethwaite, F.R.; Ratcliffe, B.; Klápště, J.; Porth, I.; Chen, C.; Stoehr, M.U.; El-Kassaby, Y.A. Genomic selection of juvenile height across a single-generational gap in Douglas-fir. *Heredity* **2019**, *122*, 848–863. [[CrossRef](#)]
20. Bhat, J.A.; Ali, S.; Salgotra, R.K.; Mir, Z.A.; Dutta, S.; Jadon, V.; Tyagi, A.; Mushtaq, M.; Jain, N.; Singh, P.K.; et al. Genomic selection in the era of next generation sequencing for complex traits in plant breeding. *Front. Genet.* **2016**, *7*, 221. [[CrossRef](#)]

21. Crossa, J.; Pérez-Rodríguez, P.; Cuevas, J.; Montesinos-López, O.; Jarquín, D.; de los Campos, G.; Burgueño, J.; Camacho-González, J.M.; Pérez-Elizalde, S.; Beyene, Y.; et al. Genomic selection in plant breeding: Methods, models, and perspectives. *Trends Plant Sci.* **2017**, *22*, 961–975. [[CrossRef](#)]
22. Resende, M.D.V.; Resende, M.F.R., Jr.; Sansaloni, C.P.; Petrolí, C.D.; Missiaggia, A.A.; Aguiar, A.M.; Abad, J.M.; Takahashi, E.K.; Rosado, A.M.; Faria, D.A.; et al. Genomic selection for growth and wood quality in *Eucalyptus*: Capturing the missing heritability and accelerating breeding for complex traits in forest trees. *New Phytol.* **2012**, *194*, 116–128. [[CrossRef](#)] [[PubMed](#)]
23. Grattapaglia, D.; Resende, M.D.V.; Resende, M.R.; Sansaloni, C.P.; Petrolí, C.D.; Missiaggia, A.A.; Takahashi, E.K.; Zamprogno, K.C.; Kilian, A. Genomic Selection for growth traits in *Eucalyptus*: Accuracy within and across breeding populations. *BMC Proc.* **2011**, *5*, O16. [[CrossRef](#)]
24. Grattapaglia, D.; Silva-Junior, O.B.; Resende, R.T.; Cappa, E.P.; Müller, B.S.F.; Tan, B.; Isik, F.; Blaise Ratcliffe, B.; Yousry, A. El-Kassaby, Y.A Quantitative genetics and genomics converge to accelerate forest tree breeding. *Front. Plant Sci.* **2018**, *9*, 1693. [[CrossRef](#)] [[PubMed](#)]
25. Iwata, H.; Hayashi, T.; Tsumura, Y. Prospects for genomic selection in conifer breeding: A simulation study of *Cryptomeria japonica*. *Tree Genet. Genomes* **2011**, *7*, 747–758. [[CrossRef](#)]
26. Resende, M.F.R., Jr.; Munoz, P.; Acosta, J.J.; Peter, G.F.; Davis, J.M.; Grattapaglia, D.; Resende, M.D.V.; Kirst, M. Accelerating the domestication of trees using genomic selection: Accuracy of prediction models across ages and environments. *New Phytol.* **2012**, *193*, 617–624. [[CrossRef](#)]
27. Cros, D.; Mbo-Nkoulou, L.; Bell, J.M.; Oum, J.; Masson, A.; Soumahoro, M.; Tran, D.M.; Achour, Z.; Le Guen, V.; Clement-Demange, A. Within-family genomic selection in rubber tree (*Hevea brasiliensis*) increases genetic gain for rubber production. *Ind. Crops Prod.* **2019**, *138*, 111464. [[CrossRef](#)]
28. Westbrook, J.W.; Zhang, Q.; Mandal, M.K.; Jenkins, E.V.; Barth, L.E.; Jenkins, J.W.; Grimwood, J.; Schmutz, J.; Holliday, J.A. Optimizing genomic selection for blight resistance in American chestnut backcross populations: A trade-off with American chestnut ancestry implies resistance is polygenic. *Evol. Appl.* **2020**, *13*, 31–47. [[CrossRef](#)]
29. Stocks, J.J.; Metheringham, C.L.; Plumb, W.J.; Lee, S.J.; Kelly, L.J.; Nichols, R.A.; Buggs, R.J.A. Genomic basis of European ash tree resistance to ash dieback fungus. *Nat. Ecol. Evol.* **2019**, *3*, 1686–1696. [[CrossRef](#)]
30. Lenz, P.R.N.; Nadeau, S.; Mottet, M.-J.; Perron, M.; Isabel, N.; Beaulieu, J.; Bousquet, J. Multi-trait genomic selection for weevil resistance, growth, and wood quality in Norway spruce. *Evol. Appl.* **2020**, *13*, 76–94. [[CrossRef](#)]
31. Beaulieu, J.; Nadeau, S.; Ding, C.; Celedon, J.M.; Azaiez, A.; Carol Ritland, C.; Laverdière, J.-P.; Marie Deslauriers, M.; Adams, G.; Fullarton, M.; et al. Genomic selection for resistance to spruce budworm in white spruce and relationships with growth and wood quality traits. *Evol. Appl.* **2020**, 1–19. [[CrossRef](#)]
32. Bouvet, J.-M.; Ekomonod, C.G.M.; Brendelf, O.; Laclaug, J.-P.; Bouillet, J.-P.; Epron, D. Selecting for water use efficiency, wood chemical traits and biomass with genomic selection in a *Eucalyptus* breeding program. *For. Ecol. Manag.* **2020**, *465*, 118092. [[CrossRef](#)]
33. Denis, M.; Bouvet, J.-M. Efficiency of genomic selection with models including dominance effect in the context of *Eucalyptus* breeding. *Tree Genet. Genomes* **2013**, *9*, 37–51. [[CrossRef](#)]
34. de Almeida Filho, J.E.; Guimarães, J.F.R.; de Silva, F.F.; de Resende, M.D.V.; Muñoz, P.; Kirst, M.; Resende, M.F.R., Jr. The contribution of dominance to phenotype prediction in a pine breeding and simulated population. *Heredity* **2016**, *117*, 33–41. [[CrossRef](#)] [[PubMed](#)]
35. Li, Y.; Dungey, H.S. Expected benefit of genomic selection over forward selection in conifer breeding and deployment. *PLoS ONE* **2018**, *13*, e0208232. [[CrossRef](#)] [[PubMed](#)]
36. Stejskal, J.; Lstibůrek, M.; Klápště, J.; Čepl, J.; El-Kassaby, Y.A. Effect of genomic prediction on response to selection in forest tree breeding. *Tree Genet. Genomes* **2018**, *14*, 74. [[CrossRef](#)]
37. Hiraoka, Y.; Fukatsu, E.; Mishima, K.; Hirao, T.; Teshima, K.M.; Tamura, M.; Tsubomura, M.; Iki, T.; Kurita, M.; Takahashi, M.; et al. Potential of genome-wide studies in unrelated plus trees of a coniferous species, *Cryptomeria japonica* (Japanese cedar). *Front. Plant Sci.* **2018**, *9*, 1322. [[CrossRef](#)]
38. Müller, B.S.F.; Neves, L.G.; de Almeida Filho, J.E.; Resende, M.F., Jr.; Muñoz, P.R.; dos Santos, P.E.T.; Filho, E.P.; Kirst, M.; Dario Grattapaglia, D. Genomic prediction in contrast to a genome-wide association study in explaining heritable variation of complex growth traits in breeding populations of *Eucalyptus*. *BMC Genom.* **2017**, *18*, 524. [[CrossRef](#)]

39. Ballesta, P.; Bush, D.; Silva, F.F.; Freddy Mora, F. Genomic predictions using low-density SNP markers, pedigree and GWAS information: A case study with the non-model species *Eucalyptus cladocalyx*. *Plants* **2020**, *9*, 99. [[CrossRef](#)]
40. Ballesta, P.; Maldonado, C.; Pérez-Rodríguez, P.; Mora, F. SNP and haplotype-based genomic selection of quantitative traits in *Eucalyptus globulus*. *Plants* **2019**, *8*, 331. [[CrossRef](#)]
41. Cappa, E.P.; El-Kassaby, Y.A.; Muñoz, F.; Garcia, M.N.; Villalba, P.V.; Klápště, J.; Poltri, S.N.M. Improving accuracy of breeding values by incorporating genomic information in spatial-competition mixed models. *Mol. Breed.* **2017**, *37*, 125. [[CrossRef](#)]
42. Resende, R.T.; Resende, M.D.V.; Silva, F.F.; Azevedo, C.F.; Takahashi, E.K.; Silva-Junior, O.B.; Grattapaglia, D. Assessing the expected response to genomic selection of individuals and families in *Eucalyptus* breeding with an additive-dominant model. *Heredity* **2017**, *119*, 245–255. [[CrossRef](#)] [[PubMed](#)]
43. Cappa, E.P.; El-Kassaby, Y.A.; Muñoz, F.; Garcia, M.N.; Villalba, P.V.; Klápště, J.; Poltri, S.N.M. Genomic-based multiple-trait evaluation in *Eucalyptus grandis* using dominant DArT markers. *Plant Sci.* **2018**, *271*, 27–33. [[CrossRef](#)] [[PubMed](#)]
44. de Moraes, B.F.X.; dos Santos, R.F.; de Lima, B.M.; Aguiar, A.M.; Missiaggia, A.A.; da Costa Dias, D.; Rezende, G.D.P.S.; Gonçalves, F.M.A.; Acosta, J.J.; Kirst, M.; et al. Genomic selection prediction models comparing sequence capture and SNP array genotyping methods. *Mol. Breed.* **2018**, *38*, 115. [[CrossRef](#)]
45. Cappa, E.P.; de Lima, B.M.; da Silva-Junior, O.B.; Garcia, C.C.; Shawn, D.; Mansfield, S.D.; Grattapaglia, D. Improving genomic prediction of growth and wood traits in *Eucalyptus* using phenotypes from non-genotyped trees by single-step GBLUP. *Plant Sci.* **2019**, *284*, 9–15. [[CrossRef](#)] [[PubMed](#)]
46. Mphahlele, M.M.; Isik, F.; Mostert-O'Neill, M.M.; Reynolds, S.M.; Hodge, G.R.; Myburg, A.A. Expected benefits of genomic selection for growth and wood quality traits in *Eucalyptus grandis*. *Tree Genet. Genomes* **2020**, *16*, 49. [[CrossRef](#)]
47. Suontama, M.; Klápště, J.; Telfer, E.; Graham, N.; Stovold, T.; Low, C.; McKinley, R.; Dungey, H. Efficiency of genomic prediction across two *Eucalyptus nitens* seed orchards with different selection histories. *Heredity* **2019**, *122*, 370–379. [[CrossRef](#)] [[PubMed](#)]
48. Thavamanikumar, S.; Arnold, R.J.; Luo, J.; Thumma, B.R. Genomic studies reveal substantial dominant effects and improved genomic predictions in an open-pollinated breeding population of *Eucalyptus pellita*. *G3 Genes Genomes Genet.* **2020**, *10*, 3751–3763. [[CrossRef](#)]
49. Rambolarimanana, T.; Ramamonjisoa, L.; Verhaegen, D.; Tsy, J.-M.L.P.; Jacquin, L.; Cao-Hamadou, T.-V.; Makouanzi, G.; Bouvet, J.-M. Performance of multi-trait genomic selection for *Eucalyptus robusta* breeding program. *Tree Genet. Genomes* **2018**, *14*, 71. [[CrossRef](#)]
50. Bouvet, J.-M.; Makouanzi, G.; Cros, D.; Vigneron, P. Modeling additive and non-additive effects in a hybrid population using genome-wide genotyping: Prediction accuracy implications. *Heredity* **2016**, *116*, 146–157. [[CrossRef](#)]
51. Tan, B.; Grattapaglia, D.; Martins, G.S.; Ferreira, K.Z.; Sundberg, B.; Ingvarsson, P.K. Evaluating the accuracy of genomic prediction of growth and wood traits in two *Eucalyptus* species and their F1 hybrids. *BMC Plant Biol.* **2017**, *17*, 110. [[CrossRef](#)]
52. Tan, B.; Grattapaglia, D.; Wu, H.X.; Ingvarsson, P.K. Genomic relationships reveal significant dominance effects for growth in hybrid *Eucalyptus*. *Plant Sci.* **2018**, *267*, 84–93. [[CrossRef](#)] [[PubMed](#)]
53. Souza, L.M.; Francisco, F.R.; Gonçalves, P.S.; Junior, E.J.S.; Le Guen, V.; Fritsche-Neto, R.; Souza, A.P. Genomic selection in rubber tree breeding: A comparison of models and methods for managing G×E interactions. *Front Plant Sci.* **2019**, *10*, 1353. [[CrossRef](#)] [[PubMed](#)]
54. Chen, Z.-Q.; Baisou, J.; Pan, J.; Karlsson, B.; Andersson, B.; Westin, J.; García-Gil, M.R.; Wu, H.X. Accuracy of genomic selection for growth and wood quality traits in two controlpollinated progeny trials using exome capture as the genotyping platform in Norway spruce. *BMC Genom.* **2018**, *19*, 946. [[CrossRef](#)]
55. Chen, Z.-Q.; Baisou, J.; Pan, J.; Westin, J.; Gil, M.R.G.; Wu, H.X. Increased prediction ability in Norway spruce trials using a marker × environment interaction and non-additive genomic selection model. *J. Hered.* **2019**, *110*, 830–843. [[CrossRef](#)]
56. Zhou, L.; Chen, Z.; Olsson, L.; Grahn, T.; Karlsson, B.; Wu, H.W.; Lundqvist, S.-O.; García-Gil, M.R. Effect of number of annual rings and tree ages on genomic predictive ability for solid wood properties of Norway spruce. *BMC Genom.* **2020**, *21*, 323. [[CrossRef](#)]

57. Beaulieu, J.; Doerksen, T.K.; MacKay, J.; Rainville, A.; Bousquet, J. Genomic selection accuracies within and between environments and small breeding groups in white spruce. *BMC Genom.* **2014**, *15*, 1048. [[CrossRef](#)]
58. Beaulieu, J.; Doerksen, T.; Clement, S.; MacKay, J.; Bousquet, J. Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. *Heredity* **2014**, *113*, 343–352. [[CrossRef](#)] [[PubMed](#)]
59. El-Dien, O.G.; Ratcliffe, B.; Klápště, J.; Chen, C.; Porth, I.; El-Kassaby, Y.A. Implementation of the realized genomic relationship matrix to open-pollinated white spruce family testing for disentangling additive from nonadditive genetic effects. *G3 Genes Genomes Genet.* **2016**, *6*, 743–753. [[CrossRef](#)]
60. Ratcliffe, B.; El-Dien, O.G.; Cappa, E.P.; Porth, I.; Klápště, J.; Chen, C.; El-Kassaby, Y.A. Single-step BLUP with varying genotyping effort in open-pollinated *Picea glauca*. *G3 Genes Genomes Genet.* **2017**, *7*, 935–942. [[CrossRef](#)] [[PubMed](#)]
61. El-Dien, O.G.; Ratcliffe, B.; Klápště, J.; Chen, C.; Porth, I.; El-Kassaby, Y.A. Multienvironment genomic variance decomposition analysis of open-pollinated Interior spruce (*Picea glauca* × *engelmannii*). *Mol. Breed.* **2018**, *38*, 26. [[CrossRef](#)] [[PubMed](#)]
62. Lenz, P.R.N.; Nadeau, S.; Azaiez, A.; Gérardi, S.; Deslauriers, M.; Perron, M.; Isabel, N.; Beaulieu, J.; Bousquet, J. Genomic prediction for hastening and improving efficiency of forward selection in conifer polycross mating designs: An example from white spruce. *Heredity* **2020**, *124*, 562–578. [[CrossRef](#)] [[PubMed](#)]
63. Thistlethwaite, F.R.; El-Dien, O.G.; Ratcliffe, B.; Klápště, J.; Porth, I.; Chen, C.; Michael, U.; Stoehr, M.U.; Ingvarsson, P.K.; El-Kassaby, Y.A. Linkage disequilibrium vs. pedigree: Genomic selection prediction accuracy in conifer species. *PLoS ONE* **2020**, *15*, e0232201. [[CrossRef](#)] [[PubMed](#)]
64. Lenz, P.R.N.; Beaulieu, J.; Mansfield, S.D.; Clément, S.; Despons, M.; Bousquet, J. Factors affecting the accuracy of genomic selection for growth and wood quality traits in an advanced-breeding population of black spruce (*Picea mariana*). *BMC Genom.* **2017**, *18*, 335. [[CrossRef](#)] [[PubMed](#)]
65. Fuentes-Utrilla, P.; Goswami, C.; Cottrell, J.E.; Pong-Wong, R.; Law, A.; A'Hara, S.W.; Lee, S.J.; Woolliams, J.A. QTL analysis and genomic selection using RADseq derived markers in Sitka spruce: The potential utility of within family data. *Tree Genet. Genomes* **2017**, *13*, 33. [[CrossRef](#)]
66. Ukrainetz, N.K.; Mansfield, S.D. Assessing the sensitivities of genomic selection for growth and wood quality traits in lodgepole pine using Bayesian models. *Tree Genet. Genomes* **2020**, *16*, 14. [[CrossRef](#)]
67. Li, Y.; Klápště, J.; Telfer, E.; Wilcox, P.; Graham, N.; Macdonald, L.; Dungey, H.S. Genomic selection for non-key traits in radiata pine when the documented pedigree is corrected using DNA marker information. *BMC Genom.* **2019**, *20*, 1026. [[CrossRef](#)]
68. Calleja-Rodriguez, A.; Pan, J.; Funda, T.; Chen, Z.-Q.; Baisson, J.; Isik, F.; Abrahamsson, S.; Wu, H.X. Genomic prediction accuracies and abilities for growth and wood quality traits of Scots pine, using genotyping-by-sequencing (GBS) data. *bioRxiv* **2019**. [[CrossRef](#)]
69. Resende, M.F.R., Jr.; Muñoz, P.; Resende, M.D.V.; Garrick, D.J.; Fernando, R.L.; Davis, J.M.; Jokela, E.J.; Martin, T.A.; Peter, G.F.; Kirst, M. Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). *Genetics* **2012**, *190*, 1503–1510. [[CrossRef](#)]
70. Muñoz, P.R.; Resende, M.F.R., Jr.; Gezan, S.A.; Resende, M.D.V.; de los Campos, G.; Kirst, M.; Huber, D.; Peter, G.F. Unraveling additive from nonadditive effects using genomic relationship matrices. *Genetics* **2014**, *198*, 1759–1768. [[CrossRef](#)]
71. Alves, F.C.; Balmant, K.M.; Resende, M.F.R.; Kirst, M.; de los Campos, G. Accelerating forest tree breeding by integrating genomic selection and greenhouse phenotyping. *Plant Genome* **2020**, e20048. [[CrossRef](#)]
72. Thistlethwaite, F.R.; Ratcliffe, B.; Klápště, J.; Porth, I.; Chen, C.; Stoehr, M.U.; El-Kassaby, Y.A. Genomic prediction accuracies in space and time for height and wood density of Douglas-fir using exome capture as the genotyping platform. *BMC Genom.* **2017**, *18*, 930. [[CrossRef](#)] [[PubMed](#)]
73. Sawitri, S.; Tani, N.; Na'iem, M.; Widiyatno, W.; Indrioko, S.; Uchiyama, K.; Suwa, R.; Siong Ng, K.K.; Lee, S.L.; Tsumura, Y. Potential of genome-wide association studies and genomic selection to improve productivity and quality of commercial timber species in tropical rainforest, a case study of *Shorea platyclados*. *Forests* **2020**, *11*, 239. [[CrossRef](#)]
74. Akdemir, D.; Isidro-Sánchez, J. Design of training populations for selective phenotyping in genomic prediction. *Sci. Rep.* **2019**, *9*, 1446. [[CrossRef](#)] [[PubMed](#)]
75. Robertsen, C.D.; Hjortshøj, R.L.; Janss, L.L. Genomic selection in cereal breeding. *Agronomy* **2019**, *9*, 95. [[CrossRef](#)]

76. Eckert, A.J.; van Heerwaarden, J.; Wegrzyn, J.L.; Nelson, C.D.; Ross-Ibarra, J.; Gonzalez-Martinez, S.C.; Neale, D.B. Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae). *Genetics* **2010**, *185*, 969–982. [[CrossRef](#)] [[PubMed](#)]
77. Pavy, N.; Gagnon, F.; Rigault, P.; Blais, S.; Deschênes, A.; Boyle, B.; Pelgas, B.; Deslauriers, M.; Clément, S.; Lavigne, P.; et al. Development of high-density SNP genotyping arrays for white spruce (*Picea glauca*) and transferability to subtropical and nordic congeners. *Mol. Ecol. Resour.* **2013**, *13*, 324–336. [[CrossRef](#)]
78. Silva-Junior, O.B.; Faria, D.A.; Grattapaglia, D. A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing of 240 *Eucalyptus* tree genomes across 12 species. *New Phytol.* **2015**, *206*, 1527–1540. [[CrossRef](#)]
79. Jaccoud, D.; Peng, K.M.; Feinstein, D.; Kilian, A. Diversity Arrays: A solid state technology for sequence information independent genotyping. *Nucleic Acids Res.* **2001**, *29*, e25. [[CrossRef](#)]
80. Heslot, N.; Rutkoski, J.; Poland, J.; Jannink, J.-L.; Sorrells, M.E. Impact of marker ascertainment bias on genomic selection accuracy and estimates of genetic diversity. *PLoS ONE* **2013**, *8*, e74612. [[CrossRef](#)]
81. Merida-Garcia, R.; Liu, G.; He, S.; Gonzalez-Dugo, V.; Dorado, G.; Garlvez, S.; Solís, I.; Zarco-Tejada, P.J.; Reif, J.C.; Hernandez, P. Genetic dissection of agronomic and quality traits based on association mapping and genomic selection approaches in durum wheat grown in Southern Spain. *PLoS ONE* **2019**, *14*, e0211718. [[CrossRef](#)]
82. Sansaloni, C.; Petrolis, C.; Jaccoud, D.; Carling, J.; Detering, F.; Grattapaglia, D.; Kilian, A. Diversity Arrays Technology (DArT) and next-generation sequencing combined: Genome-wide, high throughput, highly informative genotyping for molecular breeding of *Eucalyptus*. *BMC Proc.* **2011**, *5*, 54. [[CrossRef](#)]
83. Liu, C.; Sikumaran, S.; Jarquin, D.; Crossa, J.; Dreisigacker, S.; Sansaloni, C.; Reynolds, M. Comparison of array- and sequencing-based markers for genome-wide association mapping and genomic prediction in spring wheat. *Crop Sci.* **2020**, *60*, 211–225. [[CrossRef](#)]
84. Zimin, A.; Stevens, K.A.; Crepeau, M.W.; Holtz-Morris, A.; Koriabine, M.; Marçais, G.; Puiu, D.; Roberts, M.; Wegrzyn, J.L.; de Jong, P.J.; et al. Sequencing and assembly of the 22-Gb loblolly pine genome. *Genetics* **2014**, *196*, 875–890. [[CrossRef](#)] [[PubMed](#)]
85. Tuskan, G.A.; Difazio, S.; Jansson, S.; Bohlmann, J.; Grigoriev, I.; Hellsten, U. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **2006**, *313*, 1596–1604. [[CrossRef](#)] [[PubMed](#)]
86. Neves, L.G.; Davis, J.M.; Barbazuk, W.B.; Kirst, M. Whole-exome targeted sequencing of the uncharacterized pine genome. *Plant J.* **2013**, *75*, 146–156. [[CrossRef](#)] [[PubMed](#)]
87. Suren, H.; Hodgins, K.A.; Yeaman, S.; Nurkowski, K.A.; Smets, P.; Rieseberg, L.H.; Aitken, S.N.; Holliday, J.A. Exome capture from the spruce and pine gigagenomes. *Mol. Ecol. Res.* **2016**, *16*, 1136–1146. [[CrossRef](#)]
88. Pyhäjärvi, T.; Kujala, S.T.; Savolainen, O. 275 years of forestry meets genomics in *Pinus sylvestris*. *Evol. Appl.* **2020**, *13*, 11–30. [[CrossRef](#)]
89. Li, Y.; Suontama, M.; Burdon, R.D.; Dungey, H.S. Genotype by environment interactions in forest tree breeding: Review of methodology and perspectives on research and application. *Tree Genet. Genom.* **2017**, *13*, 60. [[CrossRef](#)]
90. Cabrera-Bosquet, L.; Crossa, J.; von Zitzewitz, J.; Maria Dolors Serret, M.D.; Araus, L.J. High-throughput phenotyping and genomic selection: The frontiers of crop breeding converge. *J. Integr. Plant Biol.* **2012**, *54*, 312–320. [[CrossRef](#)]
91. Fiorani, F.; Schurr, U. Future scenarios for plant phenotyping. *Annu. Rev. Plant Biol.* **2013**, *64*, 267–291. [[CrossRef](#)]
92. Choudhury, S.D.; Bashyam, S.; Qiu, Y.; Awada, T. Holistic and component plant phenotyping using temporal image sequence. *Plant Methods* **2018**, *14*, 35. [[CrossRef](#)] [[PubMed](#)]
93. Furbank, R.T.; Tester, M. Phenomics—Technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.* **2011**, *16*, 635–644. [[CrossRef](#)] [[PubMed](#)]
94. Available online: <https://www.quantitative-plant.org/> (accessed on 12 October 2020).
95. Wang, Y.; Wang, D.; Shi, P.; Omasa, K. Estimating rice chlorophyll content and leaf nitrogen concentration with a digital still color camera under natural light. *Plant Methods* **2014**, *10*, 36. [[CrossRef](#)] [[PubMed](#)]
96. Salekdeh, G.H.; Reynolds, M.; Bennett, J.; Boyer, J. Conceptual framework for drought phenotyping during molecular breeding. *Trends Plant Sci.* **2009**, *14*, 488–496. [[CrossRef](#)] [[PubMed](#)]

97. Rutkoski, J.; Poland, J.; Mondal, S.; Autrique, E.; González Pérez, L.; Crossa, J.; Reynolds, M.; Ravi, S. Predictor traits from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. *G3 Genes Genomes Genet.* **2016**, *6*, 2799–2808. [[CrossRef](#)]
98. Juliana, P.; Montesinos-López, O.A.; Crossa, J.; Mondal, S.; González Pérez, L.; Poland, J.; Huerta-Espino, J.; Crespo-Herrera, L.; Govindan, V.; Dreisigacker, S.; et al. Integrating genomic-enabled prediction and high-throughput phenotyping in breeding for climate-resilient bread wheat. *Theor. Appl. Genet.* **2019**, *132*, 177–194. [[CrossRef](#)] [[PubMed](#)]
99. Krause, M.R.; Mondal, S.; Crossa, J.; Singh, R.P.; Pinto, F.; Haghghattalab, A.; Shrestha, S.; Rutkoski, J.; Gore, M.A.; Sorrells, M.E.; et al. Aerial high-throughput phenotyping enabling indirect selection for grain yield at the early-generation seed-limited stages in breeding programs. *Crop Sci.* **2020**. [[CrossRef](#)]
100. Lyra, D.H.; Virlet, N.; Sadeghi-Tehran, P.; Hassall, K.L.; Wingen, L.U.; Orford, S.; Griffiths, S.; Hawkesford, M.J.; Gancho, T.; Slavov, G.T. Functional QTL mapping and genomic prediction of canopy height in wheat measured using a robotic field phenotyping platform. *J. Exp. Bot.* **2020**, *71*, 1885–1898. [[CrossRef](#)]
101. Lozada, D.N.; Godoy, J.V.; Ward, B.P.; Carter, A.H. Genomic prediction and indirect selection for grain yield in US Pacific Northwest winter wheat using spectral reflectance indices from high-throughput phenotyping. *Int. J. Mol. Sci.* **2020**, *21*, 165. [[CrossRef](#)]
102. Baba, T.; Momen, M.; Campbell, M.T.; Walia, H.; Morota, G. Multi-trait random regression models increase genomic prediction accuracy for a temporal physiological trait derived from high-throughput phenotyping. *PLoS ONE* **2020**, *15*, e0228118. [[CrossRef](#)]
103. Yonis, B.O.; del Carpio, D.P.; Wolfe, M.; Jannink, J.L.; Kulakow, P.; Ismail Rabbi, I. Improving root characterisation for genomic prediction in cassava. *Sci. Rep.* **2020**, *10*, 8003. [[CrossRef](#)] [[PubMed](#)]
104. Xu, Y.; Liu, X.; Fu, J.; Wang, H.; Wang, J.; Huang, C.; Prasanna, B.M.; Olsen, M.S.; Wang, G.; Zhang, A. Enhancing genetic gain through genomic selection: From livestock to plants. *Plant Commun.* **2020**, *1*, 100005. [[CrossRef](#)]
105. Chawade, A.; van Ham, J.; Blomquist, H.; Bagge, O.; Alexandersson, E.; Ortiz, R. High-throughput field-phenotyping tools for plant breeding and precision agriculture. *Agronomy* **2019**, *9*, 258. [[CrossRef](#)]
106. Ludovisi, R.; Tauro, F.; Salvati, R.; Khoury, S.; Mugnozza Scarascia, G.; Harfouche, A. UAV-based thermal imaging for high-throughput field phenotyping of black poplar response to drought. *Front. Plant Sci.* **2017**, *8*, 1681. [[CrossRef](#)]
107. Solvin, T.M.; Puliti, S.; Arne Steffenrem, A. Use of UAV photogrammetric data in forest genetic trials: Measuring tree height, growth, and phenology in Norway spruce (*Picea abies* L. Karst.). *Scand. J. For. Res.* **2020**, *35*, 322–333. [[CrossRef](#)]
108. Mazis, A.; Choudhury, S.D.; Morgan, P.B.; Stoerger, V.; Jeremy Hiller, J.; Ge, Y.; Awada, T. Application of high-throughput plant phenotyping for assessing biophysical traits and drought response in two oak species under controlled environment. *For. Ecol. Manag.* **2020**, *465*, 118101. [[CrossRef](#)]
109. Vidyagina, E.O.; Subbotina, N.M.; Belyi, V.A.; Lebedev, V.G.; Krutovsky, K.V.; Shestibratov, K.A. Various effects of the expression of the xyloglucanase gene from *Penicillium canescens* in transgenic aspen under semi-natural conditions. *BMC Plant Biol.* **2020**, *20*, 251. [[CrossRef](#)]
110. Martins, G.S.; Yuliarto, M.; Antes, R.; Sabki; Prasetyo, A.; Unda, F.; Mansfield, S.D.; Hodge, G.R.; Acosta, J.J. Wood and pulping properties variation of *Acacia crassicarpa* A. Cunn. ex Benth. and sampling strategies for accurate phenotyping. *Forests* **2020**, *11*, 1043. [[CrossRef](#)]
111. Lin, Z.; Hayes, B.J.; Daetwyler, H.D. Genomic selection in crops, trees and forages: A review. *Crop Pasture Sci.* **2014**, *65*, 1177–1191. [[CrossRef](#)]
112. Gao, S.; Wang, X.; Wiemann, M.C. A critical analysis of methods for rapid and nondestructive determination of wood density in standing trees. *Ann. For. Sci.* **2017**, *74*, 27. [[CrossRef](#)]
113. Schimleck, L.; Dahlen, J.; Apiolaza, L.A.; Downes, G.; Emms, G.; Evans, R.; Moore, J.; Pâques, L.; Van den Bulcke, J.; Wang, X. Non-destructive evaluation techniques and what they tell us about wood property variation. *Forests* **2019**, *10*, 728. [[CrossRef](#)]
114. White, T.L.; Adams, W.T.; Neale, D.B. *Forest Genetics*; CABI Publishing CAB International: Cambridge, UK, 2007.
115. Banerjee, B.P.; Joshi, S.; Thoday-Kennedy, E.; Pasam, R.K.; Tibbits, J.; Hayden, M.; Spangenberg, G.; Kant, S. High-throughput phenotyping using digital and hyperspectral imaging-derived biomarkers for genotypic nitrogen response. *J. Exp. Bot.* **2020**, *71*, 4604–4615. [[CrossRef](#)] [[PubMed](#)]

116. Wang, X.; Xu, Y.; Hu, Z.; Xu, C. Genomic selection methods for crop improvement: Current status and prospects. *Crop J.* **2018**, *6*, 330–340. [[CrossRef](#)]
117. Legarra, A.; Robert-Granie, C.; Croiseau, P.; Guillaume, F.; Fritz, S. Improved LASSO for genomic selection. *Genet. Res.* **2011**, *93*, 77–87. [[CrossRef](#)] [[PubMed](#)]
118. Gianola, D.; Fernandoand, R.L.; Stella, A. Genomic assisted prediction of genetic value with semi-parametric procedures. *Genetics* **2006**, *173*, 1761–1776. [[CrossRef](#)]
119. Heslot, N.; Yang, H.-P.; Sorrells, M.E.; Jannink, J.-L. Genomic selection in plant breeding: A comparison of models. *Crop Sci.* **2012**, *56*, 146–160. [[CrossRef](#)]
120. Heslot, N.; Jannink, J.-L.; Sorrells, M.E. Perspectives for genomic selection applications and research in plants. *Crop Sci.* **2015**, *55*, 1–12. [[CrossRef](#)]
121. Banta, J.A.; Richards, C.L. Quantitative epigenetics and evolution. *Heredity* **2018**, *121*, 210–224. [[CrossRef](#)]
122. Nicotra, A.B.; Atkin, O.K.; Bonser, S.P.; Davidson, A.M.; Finnegan, E.J.; Mathesius, U.; Poot, P.; Purugganan, M.D.; Richards, C.L.; Valladares, F.; et al. Plant phenotypic plasticity in a changing climate. *Trends Plant Sci.* **2010**, *15*, 684–692. [[CrossRef](#)]
123. Gugger, P.F.; Fitz-Gibbon, S.T.; Pellegrini, M.; Sork, V.L. Species-wide patterns of DNA methylation variation in *Quercus lobata* and their association with climate gradients. *Mol. Ecol.* **2016**, *25*, 1665–1680. [[CrossRef](#)]
124. Le Gac, A.-L.; Lafon-Placette, C.; Chauveau, D. Winter-dormant shoot apical meristem in poplar trees shows environmental epigenetic memory. *J. Exp. Bot.* **2018**, *69*, 4821–4837. [[CrossRef](#)]
125. Koide, Y.; Sakaguchi, S.; Uchiyama, T.; Ota, Y.; Tezuka, A.; Nagano, A.J.; Ishiguro, S.; Takamura, I.; Kishima, Y. Genetic properties responsible for the transgressive segregation of days to heading in rice. *G3 Genes Genomes Genet.* **2019**, *9*, 1655–1662. [[CrossRef](#)]
126. de Los Reyes, B.G. Genomic and epigenomic bases of transgressive segregation—New breeding paradigm for novel plant phenotypes. *Plant Sci.* **2019**, *288*, 110213. [[CrossRef](#)] [[PubMed](#)]
127. Fujimoto, R.; Uezono, K.; Ishikura, S.; Osabe, K.; Peacock, W.J.; Elizabeth, S.; Dennis, E.S. Recent research on the mechanism of heterosis is important for crop and vegetable breeding systems. *Breed. Sci.* **2018**, *68*, 145–158. [[CrossRef](#)] [[PubMed](#)]
128. Groszmann, M.; Greaves, I.K.; Fujimoto, R.; Peacock, W.J.; Dennis, E.S. The role of epigenetics in hybrid vigour. *Trends Genet.* **2013**, *29*, 684–690. [[CrossRef](#)] [[PubMed](#)]
129. Gao, M.; Huang, Q.; Chu, Y.; Ding, C.; Zhang, B.; Su, X. Analysis of the leaf methylomes of parents and their hybrids provides new insight into hybrid vigor in *Populus deltoides*. *BMC Genet.* **2014**, *15*, S8. [[CrossRef](#)]
130. Springer, N.M.; Schmitz, R.J. Exploiting induced and natural epigenetic variation for crop improvement. *Nat. Rev. Genet.* **2017**, *18*, 563–575. [[CrossRef](#)]
131. Goddard, M.E.; Emma Whitelaw, E. The use of epigenetic phenomena for the improvement of sheep and cattle. *Front. Genet.* **2014**, *5*, 247. [[CrossRef](#)]
132. Banos, D.T.; McCartney, D.L.; Patxot, M.; Anchieri, L.; Battram, T.; Christiansen, C.; Costeira, R.; Walker, R.M.; Morris, S.W.; Archie Campbell, A. Bayesian reassessment of the epigenetic architecture of complex traits. *Nat. Commun.* **2020**, *11*, 2865. [[CrossRef](#)]
133. Shah, S.; Bonder, M.J.; Marioni, R.E.; Zhu, Z.; McRae, A.F.; Zhernakova, A.; Harris, S.E.; Liewald, D.; Henders, A.K.; Mendelson, M.M.; et al. Improving phenotypic prediction by combining genetic and epigenetic associations. *Am. J. Hum. Genet.* **2015**, *97*, 75–85. [[CrossRef](#)]
134. Yang, X.; Mackenzie, S.A. Approaches to whole-genome methylome analysis in plants. In *Plant Epigenetics and Epigenomics. Methods in Molecular Biology*; Spillane, C., McKeown, P., Eds.; Humana: New York, NY, USA, 2020; Volume 2093. [[CrossRef](#)]
135. Bernardo, R.; Yu, J.M. Prospects for genomewide selection for quantitative traits in maize. *Crop Sci.* **2007**, *47*, 1082–1090. [[CrossRef](#)]
136. Wong, C.K.; Bernardo, R. Genomewide selection in oil palm: Increasing selection gain per unit time and cost with small populations. *Theor. Appl. Genet.* **2008**, *116*, 815–824. [[CrossRef](#)] [[PubMed](#)]
137. Hayes, B.; Goddard, M. Genome-wide association and genomic selection in animal breeding. *Genome* **2010**, *53*, 876–883. [[CrossRef](#)] [[PubMed](#)]
138. Sork, V.L.; Aitken, S.N.; Dyer, R.J.; Eckert, A.J.; Legendre, P.; Neale, D.B. Putting the landscape into the genomics of trees: Approaches for understanding local adaptation and population responses to changing climate. *Tree Genet. Genomes* **2013**, *9*, 901–911. [[CrossRef](#)]

139. Charlesworth, B. Fundamental concepts in genetics: Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* **2009**, *10*, 195–205. [[CrossRef](#)]
140. Grattapaglia, D. Breeding forest trees by genomic selection: Current progress and the way forward. In *Genomics of Plant Genetic Resources*; Tuberosa, R., Graner, A., Frison, E., Eds.; Springer: Dordrecht, The Netherlands, 2014; pp. 651–682.
141. Brondani, R.P.V.; Williams, E.R.; Brondani, C.; Grattapaglia, D. A microsatellite-based consensus linkage map for species of Eucalyptus and a novel set of 230 microsatellite markers for the genus. *BMC Plant Biol.* **2006**, *6*, 20. [[CrossRef](#)]
142. Neves, L.G.; Davis, J.M.; Barbazuk, W.B.; Kirst, M. A high-density gene map of loblolly pine (*Pinus taeda* L.) based on exome sequence capture genotyping. *G3 Genes Genomes Genet.* **2014**, *4*, 29–37. [[CrossRef](#)]
143. Krutovsky, K.V.; Troggio, M.; Brown, G.R.; Jermstad, K.D.; Neale, D.B. Comparative mapping in the *Pinaceae*. *Genetics* **2004**, *168*, 447–461. [[CrossRef](#)]
144. Meuwissen, T.H.E. Accuracy of breeding values of ‘unrelated’ individuals predicted by dense SNP genotyping. *Genet. Sel. Evol.* **2009**, *41*, 35–44. [[CrossRef](#)]
145. Kumar, S.; Bink, M.C.A.M.; Volz, R.K.; Bus, V.G.M.; Chagné, D. Towards genomic selection in apple (*Malus domestica* Borkh.) breeding programmes: Prospects, challenges and strategies. *Tree Genet. Genomes* **2012**, *8*, 1–14. [[CrossRef](#)]
146. Mullin, T.J.; Andersson, B.; Bastien, J.C.; Beaulieu, J.; Burdon, R.D.; Dvorak, W.S.; King, J.N.; Kondo, T.; Krakowski, J.; Lee, S.J.; et al. Economic importance, breeding objectives and achievements. In *Genetics, Genomics and Breeding of Conifers*; Plomion, C., Bousquet, J., Kole, C., Eds.; Science Publishers and CRC Press: New York, NY, USA, 2011; pp. 40–127.
147. Rajsic, P.; Weersink, A.; Navabi, A.; Peter Pauls, K. Economics of genomic selection: The role of prediction accuracy and relative genotyping costs. *Euphytica* **2016**, *210*, 259–276. [[CrossRef](#)]
148. Chang, W.-Y.; Gaston, C.; Cool, J.; Barb, R.; Thomas, B.R. A financial analysis of using improved planting stock of white spruce and lodgepole pine in Alberta, Canada: Genomic selection versus traditional breeding. *Int. J. For. Res.* **2019**, *92*, 297–310. [[CrossRef](#)]
149. Chamberland, V.; Robichaud, F.; Perron, M.; Gélinas, N.; Bousquet, J.; Beaulieu, J. Conventional versus genomic selection for white spruce improvement: A comparison of costs and benefits of plantations on Quebec public lands. *Tree Genet. Genomes* **2020**, *16*, 17. [[CrossRef](#)]
150. Park, Y.-S.; Beaulieu, J.; Bousquet, J. Multi-varietal forestry integrating genomic selection and somatic embryogenesis. In *Vegetative Propagation of Forest Trees*; Park, Y.-S., Bonga, J.M., Moon, H.-K., Eds.; National Institute of Forest Science: Seoul, Korea, 2016; pp. 302–322.
151. Salojärvi, J.; Smolander, O.-P.; Nieminen, K.; Rajaraman, S. Genome sequencing and population genomic analyses provide insights into the adaptive landscape of silver birch. *Nat. Genet.* **2017**, *49*, 904–912. [[CrossRef](#)]
152. Plomion, C.; Aury, J.; Amselem, J.; Aury, J.-M.; Leroy, T. Oak genome reveals facets of long lifespan. *Nat. Plants* **2018**, *4*, 440–452. [[CrossRef](#)] [[PubMed](#)]
153. Yasodha, R.; Vasudeva, R.; Balakrishnan, S.; Sakthi, A.R.; Abel, N.; Binai, N.; Rajashekar, B.; Bachpai, V.K.W.; Pillai, C.; Dev, S.A. Draft genome of a high value tropical timber tree, Teak (*Tectona grandis* L. f): Insights into SSR diversity, phylogeny and conservation. *DNA Res.* **2018**, *25*, 409–419. [[CrossRef](#)]
154. Wenzel, S.; Flachowsky, H.; Magda-Viola Hanke, M.-V. The Fast-track breeding approach can be improved by heat-induced expression of the FLOWERING LOCUS T genes from poplar (*Populus trichocarpa*) in apple (*Malus domestica* Borkh.). *Plant Cell Tissue Organ Cult.* **2013**, *115*, 127–137. [[CrossRef](#)]
155. Graham, T.; Scorza, R.; Wheeler, R.; Smith, B.; Dardick, C.; Dixit, A.; Raines, D.; Ann Callahan, A.; Srinivasan, C.; Spencer, L.; et al. Over-expression of *FT1* in plum (*Prunus domestica*) results in phenotypes compatible with spaceflight: A potential new candidate crop for bioregenerative life support systems. *Gravit. Space Res.* **2015**, *3*, 39–50.
156. Endo, T.; Fujii, H.; Omura, M.; Takehiko Shimada, T. Fast-track breeding system to introduce CTV resistance of trifoliate orange into citrus germplasm, by integrating early flowering transgenic plants with marker-assisted selection. *BMC Plant Biol.* **2020**, *20*, 224. [[CrossRef](#)]

157. Vicient, C.M.; Casacuberta, J.M. Impact of transposable elements on polyploid plant genomes. *Ann. Bot.* **2017**, *120*, 195–207. [[CrossRef](#)]
158. Wang, J.; Lu, N.; Yi, F.; Xiao, Y. Identification of transposable elements in conifer and their potential application in breeding. *Evol. Bioinform. Online* **2020**, *16*. [[CrossRef](#)] [[PubMed](#)]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).